

2026-1519, 2026-1520

**United States Court of Appeals
for the Federal Circuit**

GOOGLE LLC,

Appellant,

— v. —

CELLULAR SOUTH, INC.,

Appellee,

JOHN A. SQUIRES, Under Secretary of Commerce for Intellectual
Property and Director of the United States Patent and Trademark Office,

Intervenor.

*On Appeal from the United States Patent and Trademark Office,
Patent Trial and Appeal Board in Nos. IPR2025-00875
and IPR2025-00876*

BRIEF FOR APPELLANT

HEIDI L. KEEFE
MARK R. WEINSTEIN
COOLEY LLP
3175 Hanover Street
Palo Alto, California 94304
(650) 843-5000
hkeefe@cooley.com
mweinstein@cooley.com

ELIZABETH B. PRELOGAR
ELIAS S. KIM
DEV P. RANJAN
COOLEY LLP
1299 Pennsylvania Avenue,
NW,
Suite 700
Washington, DC 20004
(202) 842-7800
eprelogar@cooley.com
ekim@cooley.com
dranjan@cooley.com

Counsel for Appellant

JUNE 26, 2026

FORM 9. Certificate of Interest

Form 9 (p. 1)
March 2023

**UNITED STATES COURT OF APPEALS
FOR THE FEDERAL CIRCUIT**

CERTIFICATE OF INTEREST

Case Number 2026-1519, 2026-1520

Short Case Caption Google LLC v. Cellular South, Inc.

Filing Party/Entity Google LLC

Instructions:

1. Complete each section of the form and select none or N/A if appropriate.
2. Please enter only one item per box; attach additional pages as needed, and check the box to indicate such pages are attached.
3. In answering Sections 2 and 3, be specific as to which represented entities the answers apply; lack of specificity may result in non-compliance.
4. Please do not duplicate entries within Section 5.
5. Counsel must file an amended Certificate of Interest within seven days after any information on this form changes. Fed. Cir. R. 47.4(c).

I certify the following information and any attached sheets are accurate and complete to the best of my knowledge.

Date: 03/30/2026

Signature: /s/ Heidi L. Keefe

Name: Heidi L. Keefe

FORM 9. Certificate of Interest

Form 9 (p. 2)
March 2023

1. Represented Entities. Fed. Cir. R. 47.4(a)(1).	2. Real Party in Interest. Fed. Cir. R. 47.4(a)(2).	3. Parent Corporations and Stockholders. Fed. Cir. R. 47.4(a)(3).
Provide the full names of all entities represented by undersigned counsel in this case.	Provide the full names of all real parties in interest for the entities. Do not list the real parties if they are the same as the entities. ☒ None/Not Applicable	Provide the full names of all parent corporations for the entities and all publicly held companies that own 10% or more stock in the entities. ☐ None/Not Applicable
Google LLC		XXVI Holdings Inc.; Alphabet Inc.

☐ Additional pages attached

FORM 9. Certificate of Interest

Form 9 (p. 3)
March 2023

4. Legal Representatives. List all law firms, partners, and associates that (a) appeared for the entities in the originating court or agency or (b) are expected to appear in this court for the entities. Do not include those who have already entered an appearance in this court. Fed. Cir. R. 47.4(a)(4).

☐ None/Not Applicable ☐ Additional pages attached

Cooley LLP	Andrew Mace	Reuben Chen
Alexandra Leeper	Dev Ranjan	

5. Related Cases. Other than the originating case(s) for this case, are there related or prior cases that meet the criteria under Fed. Cir. R. 47.5(a)?

☒ Yes (file separate notice; see below) ☐ No ☐ N/A (amicus/movant)

If yes, concurrently file a separate Notice of Related Case Information that complies with Fed. Cir. R. 47.5(b). **Please do not duplicate information.** This separate Notice must only be filed with the first Certificate of Interest or, subsequently, if information changes during the pendency of the appeal. Fed. Cir. R. 47.5(b).

6. Organizational Victims and Bankruptcy Cases. Provide any information required under Fed. R. App. P. 26.1(b) (organizational victims in criminal cases) and 26.1(c) (bankruptcy case debtors and trustees). Fed. Cir. R. 47.4(a)(6).

☒ None/Not Applicable ☐ Additional pages attached

TABLE OF CONTENTS

	Page
STATEMENT OF RELATED CASES	1
JURISDICTIONAL STATEMENT	2
STATEMENT OF THE ISSUES	3
INTRODUCTION	4
STATEMENT OF THE CASE.....	6
A. Legal Background	6
B. Discretionary Denials and the Settled Expectations Rule.....	9
C. Procedural History.....	13
SUMMARY OF ARGUMENT	14
A. This Court Has Jurisdiction to Review Google’s Appeal Raising a Colorable Equal Protection Challenge to the PTO’s Reliance on the Settled Expectations Rule.....	15
B. On the Merits, the Application of the Settled Expectations Rule to Google Contravenes the Equal Protection Component of the Fifth Amendment	17
STANDARD OF REVIEW	18
ARGUMENT	19
I. This Court Has Jurisdiction to Review Google’s Appeal Raising a Colorable Equal Protection Challenge to the Director’s Reliance on the Settled Expectations Rule	19
A. Section 314(d) does not preclude this Court from reviewing an appeal of an institution decision that presents a colorable constitutional claim	20
1. Courts must interpret judicial-review bars not to apply to constitutional claims if that construction is fairly possible	20
2. Section 314(d) does not bar appellate review of colorable constitutional claims	25
3. This Court’s precedents do not foreclose this interpretation.....	28
4. The availability of mandamus does not require a different result	30

TABLE OF CONTENTS
(continued)

	Page
B. Google’s equal protection challenge is a colorable constitutional claim	32
C. This Court alternatively has jurisdiction to grant mandamus	37
II. The Application of the Settled Expectations Rule to Google Contravenes the Equal Protection Component of the Due Process Clause.....	38
A. The equal protection component of the Due Process Clause prohibits the government from applying classifications based on arbitrary or irrelevant factors.....	38
B. The application of the settled expectations rule to Google’s IPR petitions did not have a rational connection to a legitimate government objective	41
1. The settled expectations rule does not protect reliance interests or serve the purposes of a statute of limitations.....	42
2. Encouraging “early challenges” to patent validity is irrational.....	47
3. Patent age is not correlated to patent validity.....	50
4. Review of patent validity is a critical component of the patent system	53
CONCLUSION.....	57

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>Air-Way Elec. Appliance Corp. v. Day</i> , 266 U.S. 71 (1924).....	39
<i>Allied Chem. Corp. v. Daiflon, Inc.</i> , 449 U.S. 33 (1980).....	31
<i>Almond Bros. Lumber Co. v. United States</i> , 721 F.3d 1320 (Fed. Cir. 2013)	39
<i>Am. Fed’n of Gov’t Emps. Loc. 1 v. Stone</i> , 502 F.3d 1027 (9th Cir. 2007)	23
<i>Apple Inc. v. Vidal</i> , 63 F.4th 1 (Fed. Cir. 2023)	7, 8
<i>Armour v. City of Indianapolis</i> , 566 U.S. 673 (2012).....	40, 50, 52
<i>Arthrex, Inc. v. Smith & Nephew, Inc.</i> , 880 F.3d 1345 (Fed. Cir. 2018)	29
<i>Axon Enter., Inc. v. FTC</i> , 598 U.S. 175 (2023).....	26, 27
<i>Bates v. Nicholson</i> , 398 F.3d 1355 (Fed. Cir. 2005)	20, 29
<i>Belkin Int’l, Inc. v. Kappos</i> , 696 F.3d 1379 (Fed. Cir. 2012)	18
<i>Block v. Cmty. Nutrition Inst.</i> , 467 U.S. 340 (1984).....	20, 26
<i>Brecht v. Abrahamson</i> , 507 U.S. 619 (1993).....	29
<i>In re Cambridge Indus. USA Inc.</i> , 2025 WL 3526129 (Fed. Cir. Dec. 9, 2025).....	18

TABLE OF AUTHORITIES **(continued)**

	Page(s)
<i>Carrasco-Favela v. INS</i> , 563 F.2d 1220 (5th Cir. 1977)	40
<i>Celgene Corp. v. Peter</i> , 931 F.3d 1342 (Fed. Cir. 2019)	42
<i>Cheney v. U.S. Dist. Ct.</i> , 542 U.S. 367 (2004).....	16, 31
<i>City of Cleburne v. Cleburne Living Ctr.</i> , 473 U.S. 432 (1985).....	16, 33
<i>Cuozzo Speed Techs., LLC v. Lee</i> , 579 U.S. 261 (2016).....	<i>passim</i>
<i>de Magno v. United States</i> , 636 F.2d 714 (D.C. Cir. 1980).....	20
<i>DeSousa v. Reno</i> , 190 F.3d 175 (3d Cir. 1999)	40
<i>eBay Inc. v. MercExchange, L.L.C.</i> , 547 U.S. 388 (2006).....	54
<i>Elgin v. Dep’t of Treasury</i> , 567 U.S. 1 (2012).....	31
<i>Engquist v. Or. Dep’t of Agric.</i> , 553 U.S. 591 (2008).....	33, 34
<i>Francis v. INS</i> , 532 F.2d 268 (2d Cir. 1976)	40
<i>Hooper v. Bernalillo Cnty. Assessor</i> , 472 U.S. 612 (1985).....	39
<i>Incyte Corp. v. Sun Pharm. Indus., Inc.</i> , 136 F.4th 1096 (Fed. Cir. 2025)	49

TABLE OF AUTHORITIES **(continued)**

	Page(s)
<i>In re Intel Corp.</i> , 2026 WL 506045 (Fed. Cir. Feb. 24, 2026)	34, 35, 36
<i>Johnson v. Robison</i> , 415 U.S. 361 (1974).....	<i>passim</i>
<i>Judulang v. Holder</i> , 565 U.S. 42 (2011).....	34, 40
<i>Lepre v. Dep’t of Lab.</i> , 275 F.3d 59 (D.C. Cir. 2001).....	23, 24, 25
<i>Lozano v. Montoya Alvarez</i> , 572 U.S. 1 (2014).....	43
<i>Marozsan v. United States</i> , 852 F.2d 1469 (7th Cir. 1988)	23
<i>McGowan v. Maryland</i> , 366 U.S. 420 (1961).....	33
<i>In re McGrew</i> , 120 F.3d 1236 (Fed. Cir. 1997)	30
<i>MCM Portfolio LLC v. Hewlett-Packard Co.</i> , 812 F.3d 1284 (Fed. Cir. 2015)	6
<i>McNary v. Haitian Refugee Ctr., Inc.</i> , 498 U.S. 479 (1991).....	16, 30, 32
<i>Minton v. St. Bernard Par. Sch. Bd.</i> , 803 F.2d 129 (5th Cir. 1986)	39
<i>Mylan Lab’ys Ltd. v. Janssen Pharmaceutica, N.V.</i> , 989 F.3d 1375 (Fed. Cir. 2021)	<i>passim</i>
<i>Nazareth Hosp. v. Sec’y U.S. Dep’t of Health & Hum. Servs.</i> , 747 F.3d 172 (3d Cir. 2014)	40

TABLE OF AUTHORITIES **(continued)**

	Page(s)
<i>Nordlinger v. Hahn</i> , 505 U.S. 1 (1992).....	33
<i>Oil States Energy Servs., LLC v. Greene’s Energy Grp., LLC</i> , 584 U.S. 325 (2018).....	6
<i>Pro-Troll, Inc. v. Shortbus Flashers, LLC</i> , 2016 WL 11432003 (N.D. Cal. Dec. 23, 2016)	46
<i>SAS Inst., Inc. v. Iancu</i> , 584 U.S. 357 (2018).....	33
<i>Tapia-Acuna v. INS</i> , 640 F.2d 223 (9th Cir. 1981)	40
<i>Thryv, Inc. v. Click-To-Call Techs., LP</i> , 590 U.S. 45 (2020).....	9
<i>United States v. Arthrex</i> , 594 U.S. 1 (2021).....	8
<i>United States v. Sanchez-Gomez</i> , 584 U.S. 381 (2018).....	31
<i>Univ. Med. Ctr. v. Burwell</i> , 179 F. Supp. 3d 31 (D.D.C. 2016).....	40
<i>Webster v. Doe</i> , 486 U.S. 592 (1988).....	<i>passim</i>
<i>Woodward v. United States</i> , 871 F.2d 1068 (Fed. Cir. 1989)	21, 23, 24
 Statutes	
5 U.S.C.	
§ 701(a)(2)	22
§ 8128(b) (1994 ed.)	24

TABLE OF AUTHORITIES **(continued)**

	Page(s)
10 U.S.C. § 681(a)	23
28 U.S.C.	
§ 1295(a)(4)	8, 19, 28
§ 1295(a)(4)(A).....	2
§ 1361	31
§ 1651	37
35 U.S.C.	
§ 282	42
§ 286	12
§ 311	2, 7
§ 311(a).....	49
§ 311(c).....	35
§ 314	2, 27
§ 314(a).....	8, 9, 12
§ 314(b).....	7
§ 314(d).....	<i>passim</i>
§ 315(b).....	7, 35
§ 315(d).....	35
§ 315(e)	45
§ 315(e)(2)	49
§ 316(a)(11)	45
§ 316(b).....	56
§§ 321-329	7
§ 321(a).....	49
§ 325(d).....	9, 12
38 U.S.C. § 211(a)	22
50 U.S.C. § 403(c) (1988 ed.).....	23
Act of Dec. 12, 1980, Pub. L. No. 96-517, § 302, 94 Stat. 3015, 3015–16	42
America Invents Act, Pub. L. No. 112-29, 125 Stat. 284 (2011)	7
Patent Act of 1790, ch. 7, § 5, 1 Stat. 109 (1790)	42

TABLE OF AUTHORITIES

(continued)

	Page(s)
Rules and Regulations	
37 C.F.R.	
§ 42.4(a)	7
§ 42.75	8
Federal Circuit Rule 47.4(a)(5).....	1
Federal Circuit Rule 47.5(b)	1
Other Authorities	
157 Cong. Rec. S1363 (daily ed. Mar. 8, 2011) (statement of Sen. Schumer).....	44
<i>Apple, Inc. v. Fintiv, Inc.</i> , IPR2020-00019, Paper 11 (P.T.A.B. Mar. 20, 2020).....	10, 11
Brian J. Love, <i>An Empirical Study of Patent Litigation Timing: Could a Patent Term Reduction Decimate Trolls Without Harming Innovators?</i> , 161 U. Pa. L. Rev. 1309 (2013).....	52
Carl Shapiro, <i>Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting</i> , 1 Innovation Pol'y & Econ. 119 (2001).....	55
Christina Sperry, <i>Artificial Intelligence (AI) Takes a Role at the USPTO Patent Examination Process</i> , Mintz (Nov. 8, 2022).....	51
Colleen Chien, Christian Helmers & Alfred Spigarelli, <i>Inter Partes Review and the Design of Post-Grant Patent Reviews</i> , 33 Berkeley Tech. L.J. 817 (2018)	48
<i>Dabico Airport Solutions Inc. v. AXA Power ApS</i> , IPR2025-00408, Paper 21 (Director June 18, 2025)	11, 12, 42
Dennis Crouch, <i>An Era of No: The USPTO's New 0% Institution Rate</i> , Patently-O (Nov. 12, 2025)	13

TABLE OF AUTHORITIES

(continued)

	Page(s)
Fed. Trade Comm'n, <i>To Promote Innovation: The Proper Balance of Competition and Patent Law and Policy</i> 5 (Oct. 2003)	55
Forrest A. Jones, Kara A. Specht & Emma N. Ng, <i>Unpacking the PTAB's Patent Age Denial Doctrine Data</i> , Finnegan, Henderson, Farabow, Garrett & Dunner, LLP, (Aug. 29, 2025)	12
<i>Gen. Plastic Indus. Co. v. Canon Kabushiki Kaisha</i> , IPR2016-01357, -01358, -01359, -01360, -01361, Paper 19 (P.T.A.B. Sept. 6, 2017)	9
GovFacts, <i>How Long Does It Take to Get a Patent?</i> (last updated Dec. 5, 2025)	50
H.R. Rep. No. 96-1307 (1980)	6
H.R. Rep. No. 112-98, pt. 1 (2011)	<i>passim</i>
<i>iRhythm Technologies, Inc. v. Welch Allyn</i> , IPR2025-00363, -00374, -00376, -00377, -00378, Paper 10 (P.T.A.B. Jun. 6, 2025)	11
Jonathan DeFosse, Samuel Smith & Kenzo Kasai <i>Data Undermines USPTO's 'Settled Expectations' Doctrine</i> , Law360 (Aug. 29, 2025)	52
Josh Landau, <i>Inter Partes Review: Five Years, Over \$2 Billion Saved</i> , Patent Progress (Sept. 14, 2017)	44
Kenneth G. Huang et al., <i>Escaping the Patent Trolls: The Impact of Non-practicing Entity Litigation on Firm Innovation Strategies</i> , 45 Strategic Mgmt. J. 1954 (Rev. Oct. 2024)	52
Mark A. Lemley & Carl Shapiro, <i>Probabilistic Patents</i> , 19 J. Econ. Persp. 75 (2005)	45

TABLE OF AUTHORITIES **(continued)**

	Page(s)
Mark A. Lemley, <i>Rational Ignorance at the Patent Office</i> , 95 Nw. U. L. Rev. 1495 (2001)	44, 48, 54
Memorandum from Coke Morgan Stewart, Acting Under Sec'y of Com. for Intell. Prop. & Acting Dir. of the U.S. Pat. & Trademark Off., to All PTAB Judges, <i>Interim Processes for PTAB Workload Management</i> (Mar. 26, 2025).....	10
Memorandum from John A. Squires, Under Sec'y of Com. for Intell. Prop. & Dir. of the U.S. Pat. & Trademark Off., to All PTAB Judges, <i>Director Institution of AIA Trial Proceedings</i> (Oct. 17, 2025)	13
Memorandum from Scott R. Boalick, Chief Admin. Pat. Judge, to members of the Patent Trial and Appeal Board, <i>Guidance on USPTO's Rescission of "Interim Procedure for Discretionary Denials in AIA Post-Grant Proceedings with Parallel District Court Litigation"</i> (Mar. 24, 2025)	10
<i>NHK Spring Co. v. Intri-Plex Techs., Inc.</i> , IPR2018-00752, Paper 8 (P.T.A.B. Sept. 12, 2018)	10
<i>Order Denying Request for Ex Parte Reexamination</i> , Reexamination Control No. 90/015,984 (U.S. Patent No. 8,810,458) (U.S.P.T.O. May 22, 2026)	45
<i>Patent Dispute Report: 2025 in Review</i> , Unified Patents (Jan. 13, 2026)	53
<i>Perspectives on Patents: Post-Grant Review Procedures and Other Litigation Reforms</i> , Hearing Before the Subcomm. on Intellectual Property of the Comm. on the Judiciary, 109th Cong., 2d Sess. 2 (May 23, 2006) (statement of Sen. Orrin G. Hatch).....	44
Russ Meyer, <i>Motions to Stay Patent Cases Pending the Completion of Reexaminations by the Patent and Trademark Office</i> , Jackson Walker (July 6, 2009).....	46

TABLE OF AUTHORITIES **(continued)**

	Page(s)
Ryan Davis, <i>Stewart Says New Patent Policies Aim to Bring Stability</i> (Sept. 8, 2025)	42
Steve Brachmann, <i>Computer and Automotive Industries Urge SCOTUS to Eliminate the PTAB's Settled Expectations Doctrine</i> , IPWatchdog (June 1, 2026)	12
U.S. Pat. & Trademark Off., PTAB Trial Statistics: 2024 End of Year Outcome Roundup (Nov. 2024)	54
U.S. Pat. & Trademark Off., PTAB Trial Statistics: 2025 End of Year Outcome Roundup (Nov. 2025)	54
U.S. Pat. & Trademark Off., <i>U.S. Patent Activity: Calendar Years 1790 to the Present</i> , (archived Mar. 8, 2026).....	54
USPTO Designates Discretionary Considerations Decisions as Precedential and Informative, U.S. Pat. & Trademark Off. (Jan. 9, 2026)	12

STATEMENT OF RELATED CASES

Pursuant to Federal Circuit Rule 47.4(a)(5) and 47.5(b), counsel for Appellants state that there are no other appeals in or from the same civil action or proceeding in the originating tribunal that were previously before this or any other appellate court, and there are no cases known to counsel to be pending in this or any other tribunal that will directly affect or be directly affected by this Court's decision in the pending case.

JURISDICTIONAL STATEMENT

The PTO had jurisdiction over Google's petition to institute *inter partes* review under 35 U.S.C. §§ 311, 314. The PTO denied institution on October 17, 2025, Appx1, Appx8, and the Director denied review on January 12, 2026, Appx5, Appx12. This appeal was timely filed on March 11, 2026. As explained below, this Court has jurisdiction over this direct appeal under 28 U.S.C. § 1295(a)(4)(A).

STATEMENT OF THE ISSUES

1. Whether this Court has jurisdiction over a direct appeal from a decision of the PTO denying institution of an *inter partes* review where the appeal raises a colorable constitutional claim that the institution process violated the equal protection component of the Due Process Clause.
2. Whether the PTO's decision to deny institution of *inter partes* reviews for the '954 and '972 patents solely because both patents are over six years old violates the equal protection component of the Due Process Clause.

INTRODUCTION

In the America Invents Act, Congress created *inter partes* review (“IPR”) to reduce the burden of patent litigation on district courts, decrease costs for litigants, and enable the effective functioning of the patent system. The IPR system has been incredibly effective. IPR costs a fraction of the price of district court litigation; it is comparably effective and accurate; and it has saved both patent owners and challengers billions of dollars. This case concerns the arbitrary and unconstitutional unwinding of that system.

In 2025, the Patent and Trademark Office (“PTO”) debuted a new set of rules that sharply restrict access to IPR. The most prominent is the “settled expectations” rule, which effectively renders IPR unavailable for any patent that is more than six years old—in other words, the vast majority of patents. The PTO has provided almost no explanation for the rule, and what it has offered does not withstand the barest scrutiny. The effect of the settled expectations rule has been a sea change in IPR institution and a shift back to district court patent litigation—exactly what Congress created IPR to avoid.

In this appeal, Google raises a single claim: the PTO violated the equal protection component of the Fifth Amendment by applying the settled expectations rule to deny Google’s IPR petitions. The rule arbitrarily classifies

IPR petitioners based on an irrational factor: the number of years that have elapsed since issuance of the challenged patent. In this case, the PTO issued an institution order on three IPR petitions filed by Google against patents owned by the same patentee. The PTO allowed one of those IPR petitions to proceed to merits consideration where the patent issued four years ago, but it discretionarily denied the other two IPR petitions because those two patents happened to have issued six and seven years ago. No reason was given, other than the age of the patents, for this difference in treatment. Nothing in the law, the patent system, or rational governance justifies such a distinction. To the extent the settled expectations rule has any discernible objective, it is simply to end the IPR system for a substantial number of petitioners—an impermissible goal that cannot satisfy equal protection scrutiny.

This Court has jurisdiction to review Google's challenge. Although 35 U.S.C. § 314(d) bars review of some IPR institution decisions, it does not displace this Court's appellate jurisdiction to review colorable claims that an IPR institution decision contravenes the Constitution. If the rule were otherwise, the PTO could refuse to institute an IPR for unconstitutional reasons—or, as here, establish an unconstitutional institution process that classifies petitioners based on arbitrary factors—and courts would be powerless

to provide a remedy under ordinary appellate standards. That is not a tenable way to read Section 314(d).

Because this Court has jurisdiction to review the constitutionality of the settled expectations rule and that rule violates equal protection, the Court should vacate the non-institution decisions concerning Google’s IPR petitions and direct the PTO to reconsider those petitions without relying on the settled expectations rule.

STATEMENT OF THE CASE

A. Legal Background

In 1980, Congress determined that “expensive and lengthy . . . litigation” over the validity of “doubtful patents” was chilling innovation. H.R. Rep. No. 96-1307, at 3-4 (1980). Its solution was to create *ex parte* reexamination, a streamlined “administrative process[] that authorize[s] the PTO to reconsider and cancel patent claims that were wrongly issued.” *Oil States Energy Servs., LLC v. Greene’s Energy Grp., LLC*, 584 U.S. 325, 330 (2018).

Decades later, Congress determined that “questionable patents” were still “too easily obtained” and “too difficult to challenge.” *MCM Portfolio LLC v. Hewlett-Packard Co.*, 812 F.3d 1284, 1290 (Fed. Cir. 2015) (citing H.R. Rep. No. 112-98, pt. 1, at 39-40 (2011)). Accordingly, in the America Invents Act

(“AIA”), Pub. L. No. 112-29, 125 Stat. 284 (2011), Congress created an even “more efficient system for challenging patents that should not have issued.” H.R. Rep. No. 112-98, pt. 1, at 39-40 (2011). The centerpiece of that system is *inter partes* review, which allows any party to file a “petition asking the PTO” to review “whether identified claims in an issued patent comply with certain patentability requirements of novelty or obviousness over prior art.” *Apple Inc. v. Vidal*, 63 F.4th 1, 6 (Fed. Cir. 2023) (citing 35 U.S.C. § 311).

Under the AIA, a party seeking to file an IPR petition challenging the claims of a patent must comply with two timing requirements. First, the party must wait nine months after the patent’s issuance; before that point, only the “post-grant review” process is available. 35 U.S.C. §§ 321-329. Second, if the party “is served with a complaint alleging infringement of the patent,” it must file the IPR petition within one year of service. 35 U.S.C. § 315(b). The AIA imposes no other requirements related to when IPR petitions may be filed.

When a party files a petition, the AIA “assign[s] the task of determining whether to institute a review” to the PTO Director. *Apple*, 63 F.4th at 6 (citing 35 U.S.C. § 314(b)). Shortly after Congress enacted the AIA, the Director delegated his authority to make institution decisions to the Patent Trial and Appeal Board. 37 C.F.R. § 42.4(a). But after the Supreme Court’s decision in

United States v. Arthrex, 594 U.S. 1 (2021), the PTO promulgated regulations creating the “Director Review” process, which allows the Director to review any institution decision on request of a party or *sua sponte*. 37 C.F.R. § 42.75.

“For the Director to institute, certain preconditions must be met.” *Apple*, 63 F.4th at 6. In addition to ensuring that the petition satisfies the AIA’s timing requirements, the Director may not institute an IPR unless he determines “that there is a reasonable likelihood that the petitioner would prevail with respect to at least 1 of the claims challenged in the petition.” 35 U.S.C. § 314(a). The AIA does not, however, “compel[]” the Director to “institute an IPR” of every petition that “cross[es] [the] merits ‘threshold.’” *Apple*, 64 F.4th at 6.

The AIA limits appellate review of the Director’s institution decisions in certain circumstances. Specifically, while the statute gives this Court “exclusive jurisdiction . . . of an appeal from a decision of . . . the Patent Trial and Appeal Board of the [PTO] with respect to . . . inter partes review,” 28 U.S.C. § 1295(a)(4), it also provides that “[t]he determination by the Director whether to institute an inter partes review under this section shall be final and nonappealable,” 35 U.S.C. § 314(d). Accordingly, “a party generally cannot contend on appeal” that the Director’s institution decision was wrong—at least ““where the grounds for attacking the decision . . . consist of questions that are

closely tied to the application and interpretation of statutes related to the Patent Office’s decision to initiate inter partes review.’” *Thryv, Inc. v. Click-To-Call Techs., LP*, 590 U.S. 45, 53 (2020) (quoting *Cuozzo Speed Techs., LLC v. Lee*, 579 U.S. 261, 274-75 (2016)). But that general rule does not automatically apply to “appeals that implicate constitutional questions” or those “that depend on other less closely related statutes.” *Cuozzo*, 579 U.S. at 275.

B. Discretionary Denials and the Settled Expectations Rule

In the years following the AIA’s enactment, the Board primarily made institution decisions based on the merits standard set out in 35 U.S.C. § 314(a), or on other statutory factors that bear on institution decisions, *see, e.g., id.* § 325(d). The Board later began to deny IPR petitions based on various discretionary considerations not set forth in the AIA. And then it issued a series of opinions that gave these “discretionary denials” precedential force.

The Board’s first precedential discretionary denial opinion concerned “serial” IPR petitions. *See Gen. Plastic Indus. Co. v. Canon Kabushiki Kaisha*, IPR2016-01357, -01358, -01359, -01360, -01361, Paper 19 (P.T.A.B. Sept. 6, 2017). In that opinion, the Board set forth seven factors that it would consider when evaluating “follow-on” petitions submitted by a party who had already filed a petition challenging the same patent. *Id.* at 10. And in a subsequent pair

of opinions, the Board created a framework for considering whether IPR petitions should be denied based on the existence of parallel district court litigation involving the same patents. *See NHK Spring Co. v. Intrix-Plex Techs., Inc.*, IPR2018-00752, Paper 8 (P.T.A.B. Sept. 12, 2018); *Apple, Inc. v. Fintiv, Inc.*, IPR2020-00019, Paper 11 (P.T.A.B. Mar. 20, 2020).¹

On March 26, 2025, the Acting Director further elevated the role of discretionary denials by establishing a new two-phase process for institution decisions.² In the first phase, the Director (consulting with the Board) would determine whether the various discretionary denial factors weighed in favor of denial. Stewart Memo at 1. If the Director concluded that the petition did not warrant discretionary denial, then the Board would evaluate the petition based on its “merits and other non-discretionary statutory considerations.” *Id.* The

¹ Although the Director limited the application of the discretionary denial framework in a 2022 memorandum, that memorandum was rescinded in March 2025. Memorandum from Scott R. Boalick, Chief Admin. Pat. Judge, to members of the Patent Trial and Appeal Board, *Guidance on USPTO’s Rescission of “Interim Procedure for Discretionary Denials in AIA Post-Grant Proceedings with Parallel District Court Litigation”* (Mar. 24, 2025), <https://perma.cc/S6DT-TQEA>.

² Memorandum from Coke Morgan Stewart, Acting Under Sec’y of Com. for Intell. Prop. & Acting Dir. of the U.S. Pat. & Trademark Off., to All PTAB Judges, *Interim Processes for PTAB Workload Management* (Mar. 26, 2025), <https://perma.cc/JE38-Z6P5> (the “Stewart Memo”)

Acting Director also outlined new “considerations” that would influence discretionary denial decisions. *Id.* at 2. As relevant here, those considerations included what the memorandum described as “[s]ettled expectations of the parties, such as the length of time the claims ha[d] been in force.” *Id.*

A few months later, in *iRhythm Technologies, Inc. v. Welch Allyn*, IPR2025-00363, -00374, -00376, -00377, -00378, Paper 10 (P.T.A.B. Jun. 6, 2025), the Acting Director applied the settled expectations framework for the first time. *Id.* at 3. The Acting Director observed “that because one of the patents has been in force since as early as 2012 and Petitioner was aware of it as early as 2013,” “settled expectations favor[ed] denial of institution.” *Id.* Although the Acting Director found that “[s]everal arguments weigh[ed] against discretionary denial,” she declined to institute an IPR based only on the patent’s age and the owner’s “settled expectations.” *Id.* at 2-3.

Soon after, in *Dabico Airport Solutions Inc. v. AXA Power ApS*, IPR2025-00408, Paper 21 (Director June 18, 2025), the Acting Director again applied the settled expectations framework to deny institution of an IPR for an eight-year-old patent. *Id.* at 2-3. In contrast to *iRhythm*, there was no evidence that the IPR petitioner had been aware of the patent, and the patent owner never raised the settled expectations issue. *Id.*; *see generally* Patent Owner’s Br.

Requesting Discretionary Denial Under 35 U.S.C. § 325(d) and/or § 314(a), *Dabico Airport Sols. Inc. v. AXA Power ApS*, IPR2025-00408, Paper 10 (April 28, 2025). To support her conclusion that “the longer the patent has been in force, the more settled expectations should be,” the Acting Director cited the Patent Act’s six-year statute of limitations for district court infringement actions. *See Dabico*, IPR2025-00408, Paper 21, at 3 (citing 35 U.S.C. § 286).

Since *Dabico*—which the PTO has designated as “informative”—hundreds of IPR petitions have been denied with settled expectations as the only stated rationale.³ In accordance with the Acting Director’s belief that the Patent Act’s statute of limitations provides a comparator, the vast majority of IPR petitions involving patents over six years old have been discretionarily denied.⁴

In the meantime, the Director has centralized all institution decisions. On October 17, 2025, the Director issued a memorandum stating that, going

³ *See USPTO Designates Discretionary Considerations Decisions as Precedential and Informative*, U.S. Pat. & Trademark Off. (Jan. 9, 2026), <https://perma.cc/E5RE-FBJH>; Steve Brachmann, *Computer and Automotive Industries Urge SCOTUS to Eliminate the PTAB's Settled Expectations Doctrine*, IPWatchdog (June 1, 2026), <https://perma.cc/KA8C-V9DG>.

⁴ *See, e.g.*, Forrest A. Jones, Kara A. Specht & Emma N. Ng, *Unpacking the PTAB’s Patent Age Denial Doctrine Data*, Finnegan, Henderson, Farabow, Garrett & Dunner, LLP, (Aug. 29, 2025), <https://bit.ly/40UafpL>.

forward, he would personally make all IPR institution decisions.⁵ He also stated that in most cases, decisions as to institution would come via “summary notice” rather than in reasoned decisions. Squires Memo at 2. In the first three summary notices, institution was granted as to 8 petitions and denied as to 68. That marks a substantial change in the PTO’s practices: Historically the IPR institution rate had been over 65%.⁶

C. Procedural History

On May 9, 2025, Google filed IPR petitions challenging as unpatentable various claims of the ’853, ’954, and ’972 patents. Appx65, Appx233. On August 11, 2025, the Patent Owner filed requests for discretionary denial, Appx140, Appx309, which Google opposed, Appx171, Appx340.

On October 17, 2025, the Acting Chief Administrative Patent Judge issued a discretionary denial as to the ’954 and ’972 patents but referred the ’853 patent to the Board for an evaluation of the merits. Appx1, Appx8. In distinguishing between the ’954 and ’972 patents on the one hand and the ’853

⁵ Memorandum from John A. Squires, Under Sec’y of Com. for Intell. Prop. & Dir. of the U.S. Pat. & Trademark Off., to All PTAB Judges, *Director Institution of AIA Trial Proceedings* (Oct. 17, 2025), <https://perma.cc/M42H-KEYY> (the “Squires Memo”).

⁶ Dennis Crouch, *An Era of No: The USPTO’s New 0% Institution Rate*, Patently-O (Nov. 12, 2025), <https://perma.cc/UK8K-6JED>.

patent on the other, he relied entirely on the settled expectations doctrine. Appx2-3, Appx9-10. Although he acknowledged that several factors weighed against discretionary denial, he reasoned that the '954 and '972 patents had “been in force for seven and six years, respectively, creating settled expectations for [the] Patent Owner.” Appx2, Appx9. In contrast, the '853 patent had been in force for four years, which he found was not “a significant amount of time” and therefore “weigh[ed] against discretionary denial.” *Id.*

Google petitioned for Director review, explaining that the denial of its petitions based on the settled expectations rule violated federal statutes and the Equal Protection Clause. Appx198, Appx367. On January 12, 2026, the Director denied Google’s request. Appx6, Appx13.

SUMMARY OF ARGUMENT

This case presents a constitutional question that is critical to the proper functioning of the patent system: can the federal government deny access to IPR based on a wholly arbitrary factor—the age of the patent being challenged—that bears no rational connection to any legitimate government purpose? The answer is plainly no—and this Court has jurisdiction to say so.

A. This Court Has Jurisdiction to Review Google’s Appeal Raising a Colorable Equal Protection Challenge to the PTO’s Reliance on the Settled Expectations Rule

A. Section 314(d) does not preclude this Court from reviewing an appeal of a non-institution decision that presents a colorable constitutional claim. The Supreme Court and this Court have recognized a strong presumption in favor of judicial review of constitutional challenges to administrative action. Even if a statutory provision plainly precludes review of statutory claims, courts must read that provision to allow for review of constitutional claims if doing so is “fairly possible.” *Johnson v. Robison*, 415 U.S. 361, 367, 373 (1974). Here, reading Section 314(d) to allow appeals that raise only constitutional claims is not just “fairly possible”—it is the most natural way to read the provision.

Precedent leads to the same conclusion. The Supreme Court’s cases interpreting Section 314(d) hold only that it bars appeals challenging the Director’s interpretation and application of statutes governing institution. And this Court has never considered an appeal of a non-institution decision that raises colorable constitutional claims. *See, e.g., Mylan Lab ’ys Ltd. v. Janssen Pharmaceutica, N.V.*, 989 F.3d 1375, 1383 (Fed. Cir. 2021) (explaining that the appellant “fail[ed] to state a colorable claim for constitutional relief”).

Nor does the availability of mandamus mean that Section 314(d) can be read to preclude colorable constitutional claims. Mandamus does not preserve “meaningful review” of constitutional claims because it is an “extraordinary remedy” that is appropriate “only in extraordinary situations” or “exceptional circumstances.” *McNary v. Haitian Refugee Ctr., Inc.*, 498 U.S. 479, 497 (1991); *Cheney v. U.S. Dist. Ct.*, 542 U.S. 367, 380, 393 (2004).

B. Google’s appeal raises a colorable claim under the equal protection component of the Fifth Amendment. Under core equal protection principles, the government may not base differential treatment on an “arbitrary or irrational” classification that does not serve a “legitimate governmental purpose.” *City of Cleburne v. Cleburne Living Ctr.*, 473 U.S. 432, 446 (1985). An institution process that arbitrarily denies access to IPR based solely on the age of the challenged patent is no different in kind than flipping a coin to determine who can access IPR. No judicial-review bar prevents adjudication of that colorable constitutional claim.

B. On the Merits, the Application of the Settled Expectations Rule to Google Contravenes the Equal Protection Component of the Fifth Amendment

A. The equal protection component of the Due Process Clause has long been understood to prohibit government actions that are based on irrational classifications. That principle applies even when a government official acts pursuant to a statute that grants him discretion. And arbitrary classifications violate equal protection regardless of whether the challenging party had any right to the underlying benefit. No matter what is being granted or withheld, the government must—at minimum—be able to give a rational explanation for the distinctions it draws.

B. Those principles compel the conclusion that the PTO's application of the settled expectations rule to Google's IPR petitions violated equal protection. The settled expectations rule does not preserve reliance interests in patent validity or provide patent owners with repose—it simply funnels patent challenges into less efficient and far more expensive district court litigation. It will not encourage early challenges to patent validity and, if it did, the resulting flood of early challenges would drown the PTO in thousands of IPR petitions concerning patents that would never have been asserted in infringement actions. Nor does the PTO's six-year benchmark bear any relationship to the

likelihood that a patent will be found valid. In the end, the settled expectations rule broadly undermines the patent system by drastically limiting a central component of the modern patent regime—an efficient avenue to challenge invalid patents. The Court should vacate the PTO’s non-institution decision and direct the PTO to reconsider Google’s petitions without adhering to the settled expectations rule.⁷

STANDARD OF REVIEW

The PTO’s legal conclusions, statutory interpretation, and questions of jurisdiction are reviewed de novo. *Belkin Int’l, Inc. v. Kappos*, 696 F.3d 1379, 1381 (Fed. Cir. 2012).

⁷ The legality of the settled expectations framework has been challenged in multiple petitions for mandamus before this Court. *See, e.g., In re Cambridge Indus. USA Inc.*, 2025 WL 3526129 (Fed. Cir. Dec. 9, 2025) (nonprecedential). The Supreme Court recently requested a response to a petition for certiorari filed by Google challenging the PTO’s application of the settled expectations rule on statutory grounds and raising some of the jurisdictional arguments present in this appeal. *See Response Requested, Google LLC v. VirtaMove, Corp.*, No. 25-1230 (U.S. Jun. 12, 2026). As multiple amici informed the Supreme Court in urging it to grant review, the settled expectations framework is contrary to the America Invents Act and unjustifiable as a matter of policy, and Congress has not stripped this Court of jurisdiction to review those claims. Brief of Unified Patents, LLC as Amicus Curiae in Support of Petitioner, *Google LLC v. VirtaMove, Corp.*, No. 25-1230 (U.S. May 29, 2026); Brief of 35 Intellectual Property Law Professors as Amici Curiae in Support of Petitioner, *Google LLC v. VirtaMove, Corp.*, No. 25-1230 (U.S. May 29, 2026).

ARGUMENT

I. This Court Has Jurisdiction to Review Google’s Appeal Raising a Colorable Equal Protection Challenge to the Director’s Reliance on the Settled Expectations Rule

Congress vested this Court with jurisdiction to review “an appeal from a decision of . . . the Patent Trial and Appeal Board of the [PTO] with respect to . . . inter partes review.” 28 U.S.C. § 1295(a)(4). But there is an exception to that jurisdictional grant: a “determination by the Director whether to institute an inter partes review under this section shall be final and nonappealable.” 35 U.S.C. § 314(d). The Supreme Court has interpreted Section 314(d) to mean that this Court lacks jurisdiction over appeals that “attack[]” an institution decision by arguing that the PTO misapplied “statutes related to the . . . decision to initiate inter partes review.” *Cuozzo*, 579 U.S. at 274-75. But the Supreme Court has expressly stated that its holding that Section 314(d) bars review of certain statutory claims does not extend to “appeals that implicate constitutional questions.” *Id.* at 275.

This is such an appeal. Google focuses exclusively on a colorable equal protection challenge to the settled expectations rule, which subjects patent challengers to differential treatment based on the arbitrary factor of a patent-issuance date. Because Section 314(d) lacks the clear language necessary to

displace this Court’s power to review colorable constitutional claims, the Court has jurisdiction over this appeal.

A. Section 314(d) does not preclude this Court from reviewing an appeal of an institution decision that presents a colorable constitutional claim

1. Courts must interpret judicial-review bars not to apply to constitutional claims if that construction is fairly possible

The “‘strong presumption’ in favor of judicial review” serves as a background principle of interpretation for “statutes that may limit or preclude review.” *Cuozzo*, 579 U.S. at 273 (citation omitted). Under that principle, “provisions precluding judicial review” should be “give[n] a narrow construction.” *Bates v. Nicholson*, 398 F.3d 1355, 1363 (Fed. Cir. 2005); *see de Magno v. United States*, 636 F.2d 714, 721 (D.C. Cir. 1980) (“[W]here Congress has limited judicial review over questions which are within the normal competence of courts, those jurisdictional limitations are interpreted narrowly.”). Such provisions should not be interpreted to preclude review of claims that a court otherwise has jurisdiction to review unless there is “‘clear and convincing evidence’ of a contrary legislative intent.” *Block v. Cmty. Nutrition Inst.*, 467 U.S. 340, 350 (1984) (citation omitted).

That rule of interpretation applies with even greater force to constitutional claims. A statutory provision does not “preclude judicial review of constitutional claims” unless Congress’s “intent to do so” is “clear.” *Webster v. Doe*, 486 U.S. 592, 603 (1988). Courts “require this heightened showing in part to avoid the ‘serious constitutional question’ that would arise if a federal statute were construed to deny judicial forum for a colorable constitutional claim.” *Id.* Accordingly, if a statute “indicate[s] that the [government’s] discretion” to take certain actions “is generally unfettered,” “actions claimed to be based on constitutionally infirm grounds are nevertheless subject to judicial review” in most cases. *Woodward v. United States*, 871 F.2d 1068, 1072 (Fed. Cir. 1989).

Applying that rule, the Supreme Court has repeatedly held that courts should construe provisions that expressly limit judicial review of agency action to apply only to statutory claims and not to constitutional claims if that construction is “fairly possible.” *Johnson*, 415 U.S. at 367, 373. For instance, the Court held in *Johnson* that a constitutional challenge to a denial of veterans’ benefits was not barred by a provision stating that “no . . . court of the United States shall have power or jurisdiction to review” “the decisions of the Administrator [of Veterans’ Affairs] on any question of law or fact under any

law . . . providing benefits for veterans.” *Id.* at 367 (citing 38 U.S.C. § 211(a)). At the outset, the Court noted that “[p]lainly, no explicit provision . . . bars judicial consideration of . . . constitutional claims.” *Johnson*, 415 U.S. at 367. And because the provision mentioned only “decision[s] of law or fact ‘under’ [the] statute,” it could be construed as barring only claims challenging the Administrator’s “interpretation or application of a particular provision of the statute to a particular set of facts.” *Id.* The statute thus did not “provide[] the ‘clear and convincing’ evidence of congressional intent required . . . to restrict access to judicial review” of constitutional claims. *Id.* at 373-74.

Similarly, in *Webster*, the Supreme Court concluded that a judicial-review bar applicable to employment decisions by the CIA Director did not bar constitutional claims arising from those decisions. The plaintiff in *Webster* relied on the Administrative Procedure Act’s waiver of sovereign immunity to bring “a variety of statutory and constitutional claims against the Director.” 486 U.S. at 596-97. In response, the government argued that review was unavailable because the APA barred judicial review of “agency action [that] is committed to agency discretion by law,” *id.* at 597 (quoting 5 U.S.C. § 701(a)(2)), and Section 102(c) of the National Security Act (“NSA”) provided that “the Director . . . may, in his discretion, terminate the

employment of any . . . employee . . . whenever he shall deem such termination necessary or advisable in the interests of the United States,” *id.* at 594 (quoting 50 U.S.C. § 403(c) (1988 ed.)). The Court agreed in part; it explained that by giving the Director the discretion to determine what is “necessary or advisable,” the NSA barred the plaintiff from bringing an arbitrary and capricious claim or otherwise arguing that the “termination was not ‘necessary or advisable’” under the statute. *Webster*, 486 U.S. at 600, 603. But the Court concluded that nothing in the NSA displayed an intent to “preclude consideration of colorable constitution claims arising out of the actions of the Director.” *Id.* at 603.

Circuit courts across the country have similarly read judicial-review bars to apply only to statutory claims and not to constitutional claims. *See, e.g., Woodward*, 871 F.2d at 1072; *Lepre v. Dep’t of Lab.*, 275 F.3d 59, 66-67 (D.C. Cir. 2001); *Marozsan v. United States*, 852 F.2d 1469, 1476-79 (7th Cir. 1988); *Am. Fed’n of Gov’t Emps. Loc. 1 v. Stone*, 502 F.3d 1027, 1036 (9th Cir. 2007). In *Woodward*, for example, this Court held that a statute empowering the Secretary of the Navy to “‘at any time release a Reserve under his jurisdiction from active duty’” did not divest the Court of Federal Claims from reviewing constitutional challenges to termination decisions. 871 F.2d at 1071-72 (quoting 10 U.S.C. § 681(a)). Citing *Webster*’s “heightened showing”

requirement, the Court explained that, “[w]hile the statute and cited authorities indicate that the Secretary’s discretion is generally unfettered, employment actions claimed to be based on constitutionally infirm grounds are nevertheless subject to judicial review.” *Id.* at 1072 & n.4.

The D.C. Circuit’s decision in *Lepre* is also instructive. *Lepre* held that the Federal Employees Compensation Act (FECA)—which authorized the Secretary of Labor to suspend employees’ compensation in certain circumstances—did not bar an employee from arguing that the Secretary violated the Due Process Clause by suspending his disability benefits. 275 F.3d at 63-64. The key issue was how to interpret FECA’s “judicial review provision,” *id.* at 61, which provided that “[t]he action of the Secretary . . . in allowing or denying a payment under this subchapter is . . . not subject to review . . . by a court by mandamus or otherwise.” *Id.* at 64 (quoting 5 U.S.C. § 8128(b) (1994 ed.)). Noting that there must be “special clarity of congressional intent to bar federal courts from adjudicating the merits of constitutional challenges,” *id.* at 65, the court explained that Section 8128(b) did not preclude the plaintiff’s due process challenge because it “speaks only to the action of the Secretary ‘under this chapter’ and does not rule out judicial review of constitutional challenges.” *Id.* at 67.

2. Section 314(d) does not bar appellate review of colorable constitutional claims

Read against those background principles and precedents, Section 314(d) does not preclude this Court from reviewing appeals that challenge non-institution decisions solely on constitutional grounds. To start, the text of Section 314(d) says nothing about constitutional claims. Thus, “no explicit provision” of the AIA “bars judicial consideration of . . . constitutional claims” arising from non-institution decisions. *Johnson*, 415 U.S. at 367.

Indeed, Section 314(d) expressly applies only to the “determination by the Director whether to institute an inter partes review *under this section*.” 35 U.S.C. § 314(d) (emphasis added). It thus resembles the provision at issue in *Johnson*, 415 U.S. at 367, which barred review of the Administrator’s “decision[s] of law or fact ‘under’ a statute,” and the provision at issue in *Lepre*, 275 F.3d at 67, which barred review of “the action[s] of the Secretary ‘under [a] chapter.’” As discussed, such provisions apply only to claims challenging the relevant official’s “interpretation or application of a particular provision of the [relevant] statute to a particular set of facts.” *Johnson*, 415 U.S. at 367. The provisions do not bar review of claims that raise constitutional issues, which “arise under the Constitution, not under the statute.” *Id.* (citation omitted).

Given that clear text, Section 314(d) bars appellate review only of claims that challenge the Director’s “interpretation or application of” the AIA “to a particular set of facts.” *Id.* *Cuozzo* strongly supports that conclusion. As noted, the Supreme Court held in *Cuozzo* that Section 314(d) strips this Court of jurisdiction to review *some* appeals of institution decisions. 579 U.S. at 271. But, emphasizing that Section 314(d) applies only to ““determinations . . . to initiate an inter partes review *under this section*,”” *id.* at 275, the Court limited its holding to appeals “where the grounds for attacking the decision to institute inter partes review consist of questions that are closely tied to the application and interpretation of statutes related to the Patent Office’s decision to initiate inter partes review.” *Id.* at 274-75. Although the Court did not have occasion to decide whether that holding could apply to appeals that present constitutional claims, its reasoning supports a narrow construction of Section 314(d).

The “structure of the statutory scheme” further reinforces that interpretation. *Block*, 467 U.S. at 345. Congress often bars review of issues that are within an agency’s “competence and expertise” and allows review of those that are not. *Axon Enter., Inc. v. FTC*, 598 U.S. 175, 194-95 (2023) (citation omitted). In *Webster*, for instance, the Court concluded that Congress had barred review of the plaintiff’s claim that “his termination was not

‘necessary or advisable in the interests of the United States’” but not of his claim that the CIA Director discriminated against him in violation of the Fifth Amendment. 486 U.S. at 603-04. That decision made sense because the first claim fell squarely within the CIA Director’s expertise, while the second claim did not. *Id.*; *see Axon*, 598 U.S. at 194-95 (explaining that issues of “constitutional law” are typically outside an agency’s expertise). Likewise, it makes perfect sense that Congress would have barred this Court from reviewing the Director’s application of the AIA’s statutory standards to a particular set of facts—a matter squarely within his expertise—while allowing review of the Director’s potential constitutional violations.

Finally, to the extent legislative history is relevant, nothing in the AIA’s legislative history indicates that Congress intended to bar appellate review of constitutional claims. *See Johnson*, 415 U.S. at 368 (assessing whether legislative history “demonstrate[s] a congressional intention to bar judicial review even of constitutional questions”). On the contrary, the House Judiciary Committee’s report on the AIA says nothing about Section 314(d). Its section-by-section analysis of Section 314 simply states: “Inter partes review shall not be instituted unless the information in the petition and response shows that there

is a reasonable likelihood that a claim would be found invalid.” H.R. Rep. No. 112-98, pt. 1, at 75 (2011).

In sum, nothing in the text, structure, or history of Section 314(d) “provides the ‘clear and convincing’ evidence of congressional intent required . . . to restrict access to judicial review” of constitutional claims. *Johnson*, 415 U.S. at 373-74. Accordingly, this Court has jurisdiction to review appeals that exclusively raise “colorable constitutional claims arising out of the actions of the Director”—including non-institution decisions that violate equal protection principles. *Webster*, 486 U.S. at 603.

3. This Court’s precedents do not foreclose this interpretation

This Court’s decision in *Mylan*, does not require a different conclusion. In *Mylan*, the Court observed that “[a]t a first glance,” Section 1295(a)(4) is “perhaps broad enough to reach an appeal from a decision denying institution.” 989 F.3d at 1378. But applying the rule that the “‘specific governs the general,’” the Court stated that “§ 314(d), the more specific statute, dispels any such notion.” *Id.* (citation omitted). Because Section 314(d) “prevents ‘appeal’ from a decision denying institution,” the Court explained, “parties cannot make use of § 1295(a)(4)’s jurisdictional grant.” *Id.* Based on that reasoning, the

Court stated that “no statute confers jurisdiction over appeals from decisions denying institution.” *Id.* at 1379.

Although that statement could be interpreted in isolation to mean that Section 314(d) bars review of direct appeals that involve colorable constitutional claims, there is a simple reason not to read it that way: the appeal in *Mylan* did not present colorable constitutional claims. “When a prior decision does not ‘squarely address[] [an] issue,’ a court remains ‘free to address the issue on the merits’ in a subsequent case.” *Arthrex, Inc. v. Smith & Nephew, Inc.*, 880 F.3d 1345, 1349 (Fed. Cir. 2018) (quoting *Brecht v. Abrahamson*, 507 U.S. 619, 631 (1993)). In *Mylan*, this Court expressly found that the appellant “fail[ed] to state a colorable claim for constitutional relief.” 989 F.3d at 1383. Accordingly, the Court had no reason to “squarely address” whether Section 314(d) bars direct appellate review of colorable constitutional claims. *Arthrex*, 880 F.3d at 1349. And under this Court’s longstanding rule that “provisions precluding judicial review” should be “give[n] a narrow construction,” *Mylan*’s discussion of Section 314(d) should not be read to extend more broadly than was necessary to resolve the case. *Bates*, 398 F.3d at 1363.

Reading that portion of *Mylan* any other way would be inconsistent with reasoning elsewhere in the opinion. Indeed, when addressing the availability of mandamus, *Mylan* stated that “colorable constitutional claims” must remain reviewable even if “the Director’s exercise of his discretion to deny institution” cannot be reviewed. 989 F.3d at 1382. To the extent that *Mylan* can be construed as reaching a contrary conclusion as to *appeals* of colorable constitutional claims—which it does not address—it was “not necessary to the decision of the case.” *In re McGrew*, 120 F.3d 1236, 1238 (Fed. Cir. 1997). Accordingly, this Court is “free” to—and should—conclude that Section 314(d) does not bar review of colorable constitutional claims with that issue now squarely presented here. *Id.*

4. The availability of mandamus does not require a different result

Finally, the presumption of reviewability that applies to constitutional claims requires a narrow construction of Section 314(d) regardless of whether mandamus is an available option. Courts presume that statutes limiting jurisdiction do not apply to constitutional claims in a manner that would preclude “meaningful review of the type of claims raised.” *McNary*, 498 U.S. at 497 (1991). Preserving meaningful review means assuring both that there will be no jurisdictional obstacles to review and that practical considerations—

like the absence of an adequate factual record—will not render review inadequate. *See Elgin v. Dep’t of Treasury*, 567 U.S. 1, 17-21 (2012).

Allowing constitutional claims to be raised only in mandamus petitions does not preserve meaningful review. Mandamus is an “extraordinary remedy” that is appropriate “only in extraordinary situations” or “exceptional circumstances.” *Cheney*, 542 U.S. at 380, 393; *Allied Chem. Corp. v. Daiflon, Inc.*, 449 U.S. 33, 35 (1980). To be entitled to mandamus relief, a litigant must show a “clear and indisputable” “right to issuance of the writ” that goes far beyond demonstrating a meritorious constitutional claim. *Allied Chem. Corp.*, 449 U.S. at 35, 36. And even if a petitioner can establish a constitutional violation, every court has “discretionary power” not to grant mandamus relief. *United States v. Sanchez-Gomez*, 584 U.S. 381, 390 (2018).

Indeed, if mandamus were a sufficient vehicle for vindicating constitutional rights, it would almost never be necessary to construe judicial-review bars narrowly. Mandamus is nominally available in almost every scenario in which a litigant would seek relief against a federal agency. *See* 28 U.S.C. § 1361 (“The district courts shall have original jurisdiction of any action in the nature of mandamus to compel . . . any agency . . . to perform a duty owed to the plaintiff.”). For example, in *McNary*, a group of plaintiffs claimed

their due process rights were being violated by a constitutionally inadequate agency adjudication process. *See id.* at 487. The Supreme Court held that a statute limiting most challenges to those adjudications did not apply to the plaintiffs' claims, in part because the resulting process would have "foreclose[d] all forms of meaningful judicial review." *See id.* at 496. But any of the *McNary* plaintiffs could have sought mandamus relief ordering the agency to give them constitutionally adequate process. The same would be true in many other cases where courts have held that the presumption of reviewability renders jurisdiction-stripping statutes inapplicable to constitutional claims. *See, e.g., Webster*, 486 U.S. at 603. In all cases, the presumption of meaningful review implies review that is available in more than "extraordinary" and "exceptional" circumstances.

B. Google's equal protection challenge is a colorable constitutional claim

As explained in more detail below, this appeal presents a single claim: The PTO's application of the settled expectations rule violated the equal protection component of the Fifth Amendment. *See infra* at 38-57. The PTO classified Google's IPR petitions based on the age of the patents they challenged, denying the two petitions that challenged patents issued more than six years ago and advancing to the next phase the petition that challenged a

patent issued less than six years ago. By challenging that classification as arbitrary and irrational, Google has raised a colorable constitutional claim.⁸

Equal protection principles allow challenges to all “governmental classifications that ‘affect some groups of citizens differently than others.’” *Engquist v. Or. Dep’t of Agric.*, 553 U.S. 591, 601 (2008) (quoting *McGowan v. Maryland*, 366 U.S. 420, 425 (1961)). To be sure, equal protection challenges often focus on laws that warrant “heightened review because [they] jeopardize[] exercise of a fundamental right or categorize[] on the basis of an inherently suspect characteristic,” like race or gender. *Nordlinger v. Hahn*, 505 U.S. 1, 10 (1992). But equal protection principles also prohibit any classification that does not serve a “legitimate governmental purpose” because it is “arbitrary or irrational.” *City of Cleburne*, 473 U.S. at 446. Accordingly, a party can “state[] a valid claim under the Equal Protection Clause” simply by

⁸ Google additionally believes that the settled expectations rule violates federal statutes, including the AIA and the Administrative Procedure Act, and it also believes that those arguments fall outside of Section 314(d)’s review bar. *See, e.g., SAS Inst., Inc. v. Iancu*, 584 U.S. 357, 371 (2018) (stating that “[i]f a party believes the Patent Office has engaged in shenanigans by exceeding its statutory bounds, judicial review remains available consistent with the Administrative Procedure Act” (internal quotation marks omitted)). However, Google acknowledges that this Court has held that it lacks jurisdiction to review such arguments, so it notes these statutory objections for preservation purposes only.

“alleg[ing] that she ha[s] ‘been intentionally treated differently from others similarly situated and that there is no rational basis for the difference in treatment.’” *Engquist*, 553 U.S. at 601 (citation omitted).

Google’s equal protection challenge is thus a colorable constitutional claim. The PTO “arbitrarily classified” Google “as [a] member[] of an ‘identifiable group’”: IPR petitioners challenging patents issued more than six years ago. *Id.* And as explained in more detail below, there is no rational basis for treating such petitioners differently from petitioners challenging patents issued less than six years ago. *See infra* at 38-57. The PTO’s decision to deny institution solely because Google challenged a patent issued more than six years ago is akin to denying institution based on a “coin flip”—the paradigmatic example of an arbitrary classification. *See Judulang v. Holder*, 565 U.S. 42, 56 (2011).

This Court’s decision in *In re Intel Corp.*, 2026 WL 506045 (Fed. Cir. Feb. 24, 2026) (nonprecedential), is not to the contrary. The petitioner in *Intel* challenged a patent that had recently gone through an *ex parte* reexamination. *Id.* at *1. But because the petitioner raised a prior art reference that had not been presented previously, the PTO determined that the “most efficient process” would be for the petitioner to present that reference through *ex parte*

reexamination rather than IPR. *Id.* The petitioner challenged that decision in this Court, arguing that it violated equal protection principles because it had “no rational basis.” *Id.* at *2. But in doing so, the petitioner effectively challenged the Board’s application of Section 315(d), which authorizes the Board to terminate an IPR when “another proceeding or matter involving the patent is before the Office.” 35 U.S.C. § 315(d). Thus, this Court concluded that the petitioner’s challenge was not a “colorable constitutional claim” because it “amount[ed] to nothing more than [a] disagreement[] with . . . how the PTO may have applied (or not applied) certain statutes related to institution.” *Intel*, 2026 WL 506045, at *2.

Google’s equal protection challenge is fundamentally different because there is no statute related to institution that plausibly gives the PTO authority to deny institution based on the age of the relevant patent—indeed, if anything the AIA implicitly forecloses such denials. The AIA contains only two provisions that address the timing of petitions: (1) a petitioner must wait for “9 months after the grant of a patent” to file a petition, 35 U.S.C. § 311(c), and (2) a petitioner that receives a civil complaint alleging infringement of the patent must file within one year of service, *id.* § 315(b). Thus, Google’s equal protection challenge cannot be characterized as a “disagreement[]” with “how

the PTO may have applied (or not applied) certain statutes related to institution.” *Intel*, 2026 WL 506045, at *2.

Nor does it matter that the standard for evaluating Google’s equal protection challenge overlaps with the standard for evaluating an arbitrary and capricious challenge under the APA. Every equal protection claim—including those challenging suspect qualifications—could be framed as an arbitrary and capricious claim. But that does not mean every provision barring arbitrary and capricious claims also bars equal protection challenges. Indeed, in *Webster*, the plaintiff brought both types of claims after he was allegedly fired because of his sexual orientation. *Webster*, 486 U.S. at 602. Although firing him for that reason may have been arbitrary and capricious, the Supreme Court found such a claim unreviewable—while nevertheless allowing him to proceed under an equal protection theory. *Id.* at 602-04. Thus, even if a constitutional claim raises similar issues as a statutory claim, Congress’s decision to bar review of the statutory claim does not mean that it also barred review of the constitutional claim.

In sum, Google has raised a colorable claim under equal protection principles. Because Google’s appeal involves a “colorable constitutional

claim[] arising out of the” PTO’s non-institution decisions, this Court has jurisdiction to review that claim on direct appeal. *Webster*, 486 U.S. at 603.

C. This Court alternatively has jurisdiction to grant mandamus

In the event the Court concludes that review of Google’s claim is only possible by construing this appeal as a petition for mandamus, Google is still entitled to relief. *See Mylan*, 989 F.3d at 1377 & n.1 (recognizing that mandamus review is appropriate where the Court concludes direct review is unavailable and where, as here, the appellant cited the All Writs Act, 28 U.S.C. § 1651, in its Notice of Appeal and alternatively requested mandamus).

As explained further below, *infra* at 38-57, the PTO’s denial of Google’s petition had no rational relationship to a legitimate purpose, establishing a “clear and indisputable legal right” to relief. *Mylan*, 989 F.3d at 1382. Google also has no “other adequate method of obtaining” merits review of its IPR petition free from arbitrary considerations, which is the relief it seeks here. *Id.* Finally, relief is particularly “appropriate under the circumstances” because, under recent changes to the PTO’s decisional process, future IPR petitioners will not be able to challenge institution decisions on any substantive grounds because they will never know on what basis their petitions are granted or

denied. *Id.*; *see supra* at 13. Obtaining relief is therefore particularly urgent and—at the very least—appropriate.

II. The Application of the Settled Expectations Rule to Google Contravenes the Equal Protection Component of the Due Process Clause

As the Director’s order here confirms, the PTO’s settled expectations rule divides IPR petitioners into two categories: (1) those challenging patents issued more than six years ago, and (2) those challenging patents issued less than six years ago. Petitioners in the first category—like Google for the ’954 and ’972 patents—are out of luck. Even if they present compelling evidence that the challenged claims are unpatentable, the PTO will apply the settled expectations rule to deny their petitions.

That rule is a textbook violation of the Fifth Amendment. It relies on an arbitrary classification that results in differential treatment untethered from any legitimate government purpose. Accordingly, the PTO’s decision to deny Google’s petitions based on that rule must be vacated.

A. The equal protection component of the Due Process Clause prohibits the government from applying classifications based on arbitrary or irrelevant factors

The equal protection component of the Fifth Amendment prohibits government action that is “not rationally related to the . . . asserted legislative

goal.” *Hooper v. Bernalillo Cnty. Assessor*, 472 U.S. 612, 621-22 (1985); *see Almond Bros. Lumber Co. v. United States*, 721 F.3d 1320, 1328 (Fed. Cir. 2013) (classification violates equal protection “when it rests on grounds wholly irrelevant to the achievement of the State’s objective” (citation omitted)). A “classification is not based on anything having relation to the purpose for which it is made” therefore violates bedrock equal protection principles. *Air-Way Elec. Appliance Corp. v. Day*, 266 U.S. 71, 85 (1924).

Those principles hold true even where, as here, the government retains some discretion in deciding how to allocate its resources or distribute a benefit. The equal protection guarantee exists precisely to prevent the government from sorting people into winners and losers by arbitrary means; the constitutional injury is the arbitrary process, not the adverse outcome. Because “[t]he constitutional right to equal protection . . . is not predicated upon the existence of a property right,” a patent challenger is entitled to an institution process that does not discriminate based on a factor that has no rational connection to the merits of its petition or to any valid policy goal—even if the challenger does not have a property right in or any other entitlement to IPR. *Minton v. St. Bernard Par. Sch. Bd.*, 803 F.2d 129, 132 (5th Cir. 1986).

“[R]ational-basis ‘[r]eview of an equal protection claim in the context of agency action is similar to that under the APA.’” *Cooper Hosp. / Univ. Med. Ctr. v. Burwell*, 179 F. Supp. 3d 31, 47 (D.D.C. 2016) (quoting *Nazareth Hosp. v. Sec’y U.S. Dep’t of Health & Hum. Servs.*, 747 F.3d 172, 180 (3d Cir. 2014)). Accordingly, under equal protection principles, courts “must reverse an agency policy when [they] cannot discern a reason for it,” when it is “unmoored from the purposes and concerns” of the statutes the agency is administering, or when it is based on “irrelevant comparison[s].” *Judulang*, 565 U.S. at 64; *see also Francis v. INS*, 532 F.2d 268, 272 (2d Cir. 1976) (reaching similar conclusion based on equal protection violation); *Tapia-Acuna v. INS*, 640 F.2d 223 (9th Cir. 1981) (same); *Carrasco-Favela v. INS*, 563 F.2d 1220 (5th Cir. 1977) (same); *DeSousa v. Reno*, 190 F.3d 175 (3d Cir. 1999) (same). Furthermore, it must have been rational for the “governmental decisionmaker” to believe the facts that motivated the rule are actually “true.” *Armour v. City of Indianapolis*, 566 U.S. 673, 681 (2012).

Thus, if the Court cannot discern a legitimate justification for the differential treatment that occurs from application of the settled expectations rule, the rule violates the Fifth Amendment. And there is no legitimate

justification here: The PTO crafted the settled expectations rule by relying on arbitrary or irrelevant factors that cannot support the rule.

B. The application of the settled expectations rule to Google's IPR petitions did not have a rational connection to a legitimate government objective

The Director's application of the settled expectations rule to Google violates the guarantee of equal protection. As in hundreds of other cases since the settled expectations rule was first announced, the PTO's decision not to institute an IPR of the '954 and '972 patents was based entirely on the fact that those patents were issued more than six years ago. Appx2, Appx9. The decision not to discretionarily deny IPR for the '853 patent—which issued less than six years ago—confirms that rationale and is a stark illustration of the differential treatment the settled expectations rule produces.

That differential treatment serves no rational purpose. This case demonstrates how the settled expectations rule has played out in practice: The rule eliminates the availability of IPR for the vast majority of patents. In doing so, the settled expectations rule ignores the realities of the patent system and creates significant obstacles to its proper and efficient functioning. Because that is textbook arbitrary government classification, it violates equal protection principles.

1. The settled expectations rule does not protect reliance interests or serve the purposes of a statute of limitations

The PTO's primary justification for treating challengers differently based on the settled expectations rule is that after a certain period of time, patent owners assertedly begin to rely on the continued validity of their patents.⁹ In keeping with this rationale the PTO has compared the settled expectations rule to a statute of limitations. *See Dabico*, IPR2025-00408, Paper 21, at 3. Both premises are inaccurate and irrational.

To start, patent owners have never expected their patents to be safe from validity challenges. Litigation to invalidate patents has existed for as long as the patent system and has never been restricted to patents' early years. Patent Act of 1790, ch. 7, § 5, 1 Stat. 109, 111 (1790); 35 U.S.C. § 282. Agency review of patent validity has long been available as well—*ex parte* reexamination ("EPR") has existed for over four decades and is instituted at higher rates than IPR. *See* Act of Dec. 12, 1980, Pub. L. No. 96-517, § 302, 94 Stat. 3015, 3015–16. All patent owners are therefore granted patents with the understanding that the PTO “possesse[s] the authority to reexamine—and perhaps cancel—a patent claim it had previously allowed.” *Celgene Corp. v.*

⁹ Ryan Davis, *Stewart Says New Patent Policies Aim to Bring Stability* (Sept. 8, 2025), <https://tinyurl.com/yj77ud3v>.

Peter, 931 F.3d 1342, 1359 (Fed. Cir. 2019) (quoting *Cuozzo*, 579 U.S. at 267). Since Congress has made patents subject to challenge throughout their term, protecting reliance in a patent at an earlier point is not a legitimate goal—Congress has already made the policy judgment.

Even if protecting patent owners’ reliance interests were a legitimate goal, moreover, the settled expectations rule does not achieve that goal. In fact, the rule serves none of the functions that a statute of limitations ordinarily would. Statutes of limitations “embody a ‘policy of repose, designed to protect defendants’” by “foster[ing] the ‘elimination of stale claims, and certainty about a plaintiff’s opportunity for recovery and a defendant’s potential liabilities.’” *Lozano v. Montoya Alvarez*, 572 U.S. 1, 14 (2014) (citation omitted). The settled expectations rule does not serve those goals because it has *no effect* on the availability of validity challenges in litigation. Parties may continue to file suit to challenge validity at any point over the life of the patent.

Accordingly, the primary effect of the settled expectations rule is to shift validity disputes to costly, less efficient forums—not to provide patent owners with repose. Start with district court litigation. District courts are less efficient forums for litigating patent validity because district court judges and lay juries lack the technical and legal expertise of the Board, which is “institutionally

better equipped than the Federal Courts to resolve highly technical disputes regarding patentability.” *Perspectives on Patents: Post-Grant Review Procedures and Other Litigation Reforms*, Hearing Before the Subcomm. on Intellectual Property of the Comm. on the Judiciary, 109th Cong., 2d Sess. 2 (May 23, 2006) (statement of Sen. Orrin G. Hatch). Even more importantly, district court litigation is far more expensive than IPR—sometimes ten times as costly.¹⁰ Before the AIA, just the legal fees for patent litigation were estimated to total \$2.1 billion per year. Mark A. Lemley, *Rational Ignorance at the Patent Office*, 95 Nw. U. L. Rev. 1495, 1502 (2001) (Lemley, *Rational Ignorance*). Indeed, the IPR system was created specifically to avoid the “substantial burden on U.S. courts and the U.S. economy” from “[l]itigation over invalid patents.” 157 Cong. Rec. S1363 (daily ed. Mar. 8, 2011) (statement of Sen. Schumer); *see also* H.R. Rep. No. 112–98, pt. 1, at 48 (2011) (stating that the AIA was intended to provide “quick and cost effective alternatives to litigation”). Forcing parties back into court is therefore contrary to the purpose of the AIA, reviving the very problems that Congress was attempting to solve.

¹⁰ See Josh Landau, *Inter Partes Review: Five Years, Over \$2 Billion Saved*, Patent Progress (Sept. 14, 2017), <https://perma.cc/7ATK-BXNP>.

If the settled expectations rule pushes patent challenges into EPR instead of IPR, that too will lead to inefficiencies Congress intended the AIA to avoid.¹¹ The EPR system suffers from its own problems that make it an inferior forum to IPR. EPR is non-adversarial, and so it does not achieve the same goals as an adversarial process like IPR. H.R. Rep. No. 112-98, pt. 1, at 44–46 (2011) (stating that one of EPR’s “limitations” is that “the third-party challenger ha[s] no role once the proceeding [is] initiated”). And because a finding of validity in an EPR has no estoppel effect in district court, it does little to prevent costly patent litigation. *Compare* 35 U.S.C. § 315(e), with *id.* §§ 302-307. That is especially true because EPRs are subject to no time restrictions (and often take years to complete). Mark A. Lemley & Carl Shapiro, *Probabilistic Patents*, 19 J. Econ. Persp. 75, 79 (2005); *compare* 35 U.S.C. § 316(a)(11) (stating that IPR generally must be completed within one year). As a result, courts often decline to stay ongoing patent litigation while EPRs are pending, both because doing so would allow challengers to “effectively press[] the pause button on the

¹¹ Because the PTO has begun using discretionary denials of IPR to subsequently deny EPR institution, district court litigation might be the only option for many challengers. *See, e.g., Order Denying Request for Ex Parte Reexamination*, Reexamination Control No. 90/015,984 (U.S. Patent No. 8,810,458) (U.S.P.T.O. May 22, 2026) (denying EPR because arguments had been “previously presented” in an IPR petition that was discretionarily denied).

litigation while hardly needing to lift a finger” and because EPRs are not ““guaranteed to finally resolve . . . [any] issues of validity.”” *Pro-Troll, Inc. v. Shortbus Flashers, LLC*, 2016 WL 11432003, at *1 (N.D. Cal. Dec. 23, 2016) (citation omitted).¹²

The settled expectations rule therefore creates no “repose”—it merely funnels challenges into demonstrably inferior forums that the AIA was enacted specifically to avoid. *See, e.g.*, H.R. Rep. No. 112–98, pt. 1, at 48 (2011). Contrary to statements by PTO leadership, the rule does not encourage a “stable patent system” or a “stable economy.” Davis, *supra*, at 1. Instead, it directly contravenes the AIA’s goals. By cutting off the most efficient avenue for patent challenges while leaving other mechanisms in place to attack validity, the settled expectations rule multiplies both the costs and the efforts of patent litigation, likely draining billions of dollars from the economy for whatever years it is in force. Landau, *supra*. The classification—and its result—is therefore both arbitrary and untenable.

¹² *See also* Russ Meyer, *Motions to Stay Patent Cases Pending the Completion of Reexaminations by the Patent and Trademark Office*, Jackson Walker (July 6, 2009), <https://perma.cc/76ZU-M8HN> (stating that “courts in the Eastern District of Texas have denied motions to stay pending reexamination more often than they have granted them”).

2. Encouraging “early challenges” to patent validity is irrational

PTO leadership has also asserted that treating patent challengers differently based on the settled expectations rule is beneficial because it encourages “early challenges” to patent validity. Davis, *supra*, at 2. But the factual premises of this rationale do not withstand scrutiny.

As an initial matter, encouraging early challenges to patent validity makes no sense as a matter of economics. Under the current system, parties can avoid patent validity disputes until after they are sued, or at a minimum, until the possibility of infringement litigation is concrete. This allows them to allocate resources to challenging only the highest priority patents and avoid flooding the PTAB with unnecessary challenges. By contrast, a rule requiring parties to challenge validity “early” would impose tremendous costs on companies, patent owners, and the system as a whole. Under that rule, a company would need to monitor every relevant patent granted in its field, determine which of those patents might one day be invoked in infringement suits against current or future products, and then file preemptive IPR petitions before litigation is even on the horizon. If the settled expectations rule succeeded in prompting such preemptive petitions, the PTO would be swamped with more IPR challenges than it could possibly adjudicate.

Accordingly, the primary effect of a rule designed to encourage “early” patent challenges would be to require the expenditure of significant resources on adjudicating the validity of patents that would never be litigated. “[S]ignificantly more than half of all issued patents are never used for any purpose whatsoever.” Lemley, *Rational Ignorance*, *supra*, at 1514. “Of the roughly two million patents currently in force,” only about “2,000 different patents” per year are asserted in patent litigation. *Id.* at 1501. It thus does not “matter much whether these patents are ‘really’ valid in any objective sense” and it makes no sense to spend money adjudicating that question. *Id.* at 1514. And although IPRs are substantially less expensive than district court litigation, the typical IPR still costs an average of nearly \$300,000 to litigate through completion. Colleen Chien, Christian Helmers & Alfred Spigarelli, *Inter Partes Review and the Design of Post-Grant Patent Reviews*, 33 Berkeley Tech. L.J. 817, 837 (2018). If hundreds of thousands of new IPR petitions were filed each year to avoid application of the settled expectations rule, the cost would run in the billions of dollars per year.

Even if the costs were not prohibitive, the legal consequence of raising such “early challenges” likely would deter preemptive IPR petitions. In many cases, patent challengers would be filing challenges years before even making

a product that potentially infringes the patent in question. Although the lack of a concrete dispute would not prevent challengers from participating in an IPR, *see* 35 U.S.C. §§ 311(a), 321(a), Article III standing rules would prevent them from appealing an adverse decision to this Court. *See Incyte Corp. v. Sun Pharm. Indus., Inc.*, 136 F.4th 1096, 1104-05 (Fed. Cir. 2025). At the same time, the challengers might be estopped from raising any arguments concerning the patent’s validity on which they did not prevail in an IPR in any future district court litigation. *See* 35 U.S.C. § 315(e)(2). Given those adverse incentives, it is unlikely that prospective patent challengers would take up the invitation to file a premature IPR petition to avoid the settled expectations rule.

At bottom, the PTO’s focus on incentivizing “early challenges” rests on highly suspect factual premises. The increased costs and legal consequences of bringing such challenges will likely deter parties from bringing them at all. And if the settled expectations rule does result in such challenges, it could grind the patent system to a halt. Meanwhile, patent owners will be incentivized to delay filing suit until they are assured that IPR will be unavailable. That will result in even less efficient patent litigation in courts, as parties will have to reckon with fading memories and the increasing difficulty of finding invalidating prior art. In sum, the premise the PTO has offered for the settled

expectations rule blinks reality and cannot “rationally . . . be[] considered to be true.” *Armour*, 566 U.S. at 681.

3. Patent age is not correlated to patent validity

To the extent the PTO believes that patent age is a proxy for patent validity, that rationale for differential treatment of challengers also fails. To start, the issuance date of a patent is in many instances arbitrary. After a patent application is filed, the amount of time before the patent is issued can vary from less than a year to over a decade or more.¹³ The length of prosecution for a patent application can vary dramatically based on factors that have nothing to do with the merits of the underlying invention. For example, differences in examiner experience, technical competence, and workload can significantly influence how quickly examiners process applications and identify issues, and actions by examiners ultimately determine whether lengthy appeals are required. The length of prosecution is further varied by applicant-driven choices such as the number and scope of the claims pursued and the applicant’s willingness to amend claims during prosecution in response to office actions, all of which can further extend or compress the timeline. Any given application

¹³ GovFacts, *How Long Does It Take to Get a Patent?* (last updated Dec. 5, 2025), <https://perma.cc/3PST-NCUC>.

can thus reach allowance relatively quickly or remain in prolonged examination for years, often for administrative and behavioral factors having nothing to do with the merits of the underlying invention.

If anything, all things being equal, older patents may be *more* likely to be invalid than newer ones. For example, in technological fields that have not been the subject of extensive patenting efforts, the PTO's lack of familiarity with prior art in that field may result in a less thorough examination as compared to patent applications filed after the PTO has obtained experience in that field. The PTO has also improved the process of searching for and locating prior art, which has improved the quality of patent examination. But under the settled expectations rule, most patents examined before the implementation of modern examination tools—such as Patents End-to-End (PE2E) and AI-enhanced prior art search introduced in 2021—will be immune from IPR.¹⁴ It is thus unsurprising that before the settled expectations rule was adopted, older patents were no more likely to be upheld.¹⁵ Indeed, the PTO's own data shows

¹⁴ Christina Sperry, *Artificial Intelligence (AI) Takes a Role at the USPTO Patent Examination Process*, Mintz (Nov. 8, 2022), <https://perma.cc/N7YB-7V9S>.

¹⁵ Jonathan DeFosse, Samuel Smith & Kenzo Kasai *Data Undermines USPTO's 'Settled Expectations' Doctrine*, Law360 (Aug. 29, 2025), <https://tinyurl.com/yubhpd68>.

that the average age of a patent challenged in an IPR petition was historically over six years, 82% of the petitions challenging such patents resulted in institution, and 82% of the IPR proceedings against such patents that reached final written decision resulted in the cancellation of at least one claim. *Id.* Nor was there any bias against older patents in previous institution decisions—institution rates for IPR petitions against over-six-year-old patents were historically slightly higher than against younger patents. *Id.*

Older patents are also the ones most likely to be used in litigation by patent assertion entities (“PAEs”) and non-practicing entities (“NPEs”). Brian J. Love, *An Empirical Study of Patent Litigation Timing: Could a Patent Term Reduction Decimate Trolls Without Harming Innovators?*, 161 U. Pa. L. Rev. 1309, 1340 (2013) (stating that NPEs are responsible for 92% of infringement suits brought “late in the patent term, precisely when litigation seems most abusive”). Even though challengers have historically been successful in petitioning for IPR in response to NPE suits, NPE litigation still costs the economy hundreds of millions of dollars per year. Kenneth G. Huang et al., *Escaping the Patent Trolls: The Impact of Non-practicing Entity Litigation on Firm Innovation Strategies*, 45 Strategic Mgmt. J. 1954 (Rev. Oct. 2024).

Without IPRs, those costs will be multiplied.¹⁶ And since the adoption of the settled expectations rule, institution rates against NPEs are lower than against operating companies for the first time, as operating companies have always tended to assert patents early in the patent term. *See Love, supra*, at 1331.

Far from excluding the patents most likely to be valid from IPR, the settled expectations rule erects barriers to challenging the patents where IPR would be most likely to be successful. The result is an arbitrary classification that will produce billions of dollars in unnecessary patent litigation costs.

4. Review of patent validity is a critical component of the patent system

The ultimate effect of the settled expectations rule will be to cut off the availability of an efficient means to challenge patent validity for a class of challengers based on the happenstance of patent-issuance date. But attempting to achieve that end is irrational. Efficient review of previously granted patents is a critical component of the patent system. Even if the settled expectations rule provides some benefit to some patent owners—most notably NPEs—its adverse effects on the patent system and the economy vastly outweighs those benefits.

¹⁶ *Patent Dispute Report: 2025 in Review*, Unified Patents (Jan. 13, 2026), <https://perma.cc/8EZD-BMTE>.

Cutting off the most efficient avenue for challenges to patent validity is irrational first and foremost because extraneous, invalid patents harm the very interests the patent system is intended to advance. As commentators, courts, and Congress have all acknowledged, the bar to obtain a patent is low. H.R. Rep. No. 112-98, pt. 1, at 39 (2011) (stating that “questionable patents” are “easily obtained”). There are hundreds of thousands of patent applications filed every year, and examiners spend only 18 hours on average on an application. *See* Lemley, *Rational Ignorance*, *supra*, at 1499-1500. Patents are granted at a rate of around 300,000 per year.¹⁷ The result is “that there are too many bad patents in the system.” Lemley, *Rational Ignorance*, *supra*, at 1508; *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388, 397 (2006) (noting the existence of patents of “suspect validity”). Indeed, when patents are challenged, a substantial proportion are invalidated—depending on the year, the invalidation rate can be as high as 40%.¹⁸ And the invalidation rate would presumably be

¹⁷ U.S. Pat. & Trademark Off., *U.S. Patent Activity: Calendar Years 1790 to the Present*, (archived Mar. 8, 2026), <https://tinyurl.com/yckyz5yk>.

¹⁸ *See, e.g.*, U.S. Pat. & Trademark Off., PTAB Trial Statistics: 2025 End of Year Outcome Roundup (Nov. 2025), <https://perma.cc/GP2L-EPL4>; U.S. Pat. & Trademark Off., PTAB Trial Statistics: 2024 End of Year Outcome Roundup (Nov. 2024), <https://perma.cc/8SPJ-9KV5>.

even higher for patents that have never been asserted and were not designed to withstand the scrutiny of an adversarial challenge to validity.

Invalid patents can have serious economic consequences because “[q]uestionable patents” “deter or raise the cost of innovation” in many industries.¹⁹ In some cases, superfluous patents create a “dense web of overlapping intellectual property rights” that “have the perverse effect of stifling, not encouraging, innovation.” Carl Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting*, 1 Innovation Pol’y & Econ. 119, 120 (2001). For example, before the AIA, firms in many industries sought to “avoid infringing questionable patents” for fear of expensive patent litigation, and they would “therefore . . . refrain from entering or continuing with a particular field of research that such patents appear to cover.” FTC Report, *supra*, at 5. That result is antithetical to the core principles the patent system is intended to preserve.

The patent system therefore relies on the availability of avenues to efficiently reevaluate patents after they are granted. “[I]nter partes review helps protect the public’s ‘paramount interest in seeing that patent monopolies . . .

¹⁹ Fed. Trade Comm’n, *To Promote Innovation: The Proper Balance of Competition and Patent Law and Policy* 5 (Oct. 2003), <https://perma.cc/5XQ3-22P2> (FTC Report).

are kept within their legitimate scope.’” *Cuozzo*, 579 U.S. at 279-80 (citation omitted). If avenues like IPR were not available, the results would undermine the “integrity of the patent system” and the “economy”—in direct contravention of the AIA’s goals. *See* § 316(b).

* * *

Equal protection principles exist to guard against arbitrary government action like the settled expectations rule. When a rule produces differential treatment—when one petitioner receives access to a congressionally created review system and another is turned away—the Constitution demands that the line drawn between them correspond to a rational and legitimate governmental purpose. The settled expectations rule flunks that test. There is no principled distinction between a rule that denies IPR based on patent age and a rule that denies IPR based on a coin flip: In both instances, the line drawn consigns IPR petitioners to the reject pile based on a factor that has nothing to do with the merits of their challenges, the validity of the patents they seek to test, or any legitimate government policy. A government that can deny access to its own adjudicatory processes on those grounds is a government exercising power untethered from reason—and the Constitution does not permit it.

CONCLUSION

The Court should vacate the PTO's decisions declining to institute IPR as to the '954 and '972 patents and direct the PTO to reconsider Google's petitions without adhering to the settled expectations rule.

Dated: June 26, 2026

Respectfully submitted,

/s/ Elizabeth B. Prelogar

ADDENDUM

ADDENDUM**TABLE OF CONTENTS**

Paper/ Exhibit No.	Date	Description	Appendix Page No(s).
10	Oct. 17, 2025	Final Written Decision Denying Institution of <i>Inter Partes</i> Review ('972), IPR2025-00875	APPX001
13	Jan. 12, 2026	Order Denying Request for Director Review ('972), IPR2025-00875	APPX005
10	Oct. 17, 2025	Final Written Decision Denying Institution of <i>Inter Partes</i> Review ('954), IPR2025-00876	APPX008
13	Jan. 12, 2026	Order Denying Request for Director Review ('954), IPR2025-00876	APPX012
1001		U.S. Patent No. 9,940,972 B2 (IPR2025-00875)	APPX015
1001		U.S. Patent No. 10,218, 954 B2 (IPR2025-00876)	APPX027

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE OFFICE OF THE UNDER SECRETARY OF COMMERCE
FOR INTELLECTUAL PROPERTY AND DIRECTOR OF THE
UNITED STATES PATENT AND TRADEMARK OFFICE

GOOGLE LLC,
Petitioner,

v.

CELLULAR SOUTH, INC.,
Patent Owner.

IPR2025-00875 (Patent 9,940,972 B2)
IPR2025-00876 (Patent 10,218,954 B2)
IPR2025-00877 (Patent 11,126,853 B2)

Before KALYAN K. DESHPANDE,¹ *Acting Chief Administrative Patent Judge.*

DECISION

Denying Institution of *Inter Partes* Review in IPR2025-00875 and
IPR2025-00876 and Referring the Petition to the Board in IPR2025-00877

¹ Coke Morgan Stewart, Deputy Under Secretary of Commerce for Intellectual Property and Deputy Director of the United States Patent and Trademark Office, is recused and took no part in this decision. See <https://www.uspto.gov/sites/default/files/documents/dsco-delegation.pdf>.

IPR2025-00875 (Patent 9,940,972 B2)
IPR2025-00876 (Patent 10,218,954 B2)
IPR2025-00877 (Patent 11,126,853 B2)

Cellular South, Inc. (“Patent Owner”) filed a request for discretionary denial (Paper 6, “DD Req.”) in the above-captioned cases, and Google LLC (“Petitioner”) filed an opposition (Paper 7, “DD Opp.”).²

After considering the parties’ arguments and the record, and in view of all relevant considerations, discretionary denial of institution is appropriate in IPR2025-00875 and IPR2025-00876, but is not appropriate in IPR2025-00877. This determination is based on the totality of the evidence and arguments the parties have presented.

Some factors counsel against discretionary denial. For example, the projected final written decision due date is December 16, 2026. DD Req. 9–10. The district court trial is scheduled to begin March 15, 2027. *Id.* at 6; DD Opp. 4. As such, it is likely that a final written decision in this proceeding will issue before the district court trial occurs, reducing the risk of duplication of efforts and inconsistent outcomes. Additionally, the patent challenged in IPR2025-00877 has not been in force for a significant amount of time (issued in 2021). Early challenges favor robust, predictable patent rights and weigh against discretionary denial.

The patents challenged in IPR2025-00875 and IPR2025-00876 present different circumstances. In particular, these challenged patents have been in force for seven and six years, respectively, creating settled expectations for Patent Owner, and Petitioner does not provide persuasive reasoning why an *inter partes* review is an appropriate use of Board resources. *Dabico Airport Sols. Inc. v. AXA Power ApS*, IPR2025-00408,

² Citations are to papers in IPR2025-00875. The parties filed similar papers in IPR2025-00876 and IPR2025-00877.

IPR2025-00875 (Patent 9,940,972 B2)
IPR2025-00876 (Patent 10,218,954 B2)
IPR2025-00877 (Patent 11,126,853 B2)

Paper 21 at 2–3 (Director June 18, 2025). In the absence of any such information, the Office is disinclined to disturb the settled expectations of Patent Owner.

Although certain arguments are highlighted above, the determinations in this Decision are based on a holistic assessment of all of the evidence and arguments presented. Accordingly, the Petitions in IPR2025-00875 and IPR2025-00876 are denied under 35 U.S.C. § 314(a), and the Petition in IPR2025-00877 is referred to the Board to handle the case in the normal course, including by issuing a decision on institution addressing the merits and other non-discretionary considerations, as appropriate.

In consideration of the foregoing, it is:

ORDERED that Patent Owner’s requests for discretionary denial in IPR2025-00875 and IPR2025-00876 are *granted*;

FURTHER ORDERED that the Petitions in IPR2025-00875 and IPR2025-00876 are *denied*, and no trial is instituted;

FURTHER ORDERED that Patent Owner’s request for discretionary denial in IPR2025-00877 is *denied*;

FURTHER ORDERED that the Petition in IPR2025-00877 is referred to the Board; and

FURTHER ORDERED that neither party shall file a request for rehearing or Director Review of the decision to deny Patent Owner’s request for discretionary denial in IPR2025-00877 until the Board issues a decision on institution.

IPR2025-00875 (Patent 9,940,972 B2)

IPR2025-00876 (Patent 10,218,954 B2)

IPR2025-00877 (Patent 11,126,853 B2)

FOR PETITIONER:

Heidi Keefe

Andrew Mace

COOLEY LLP

hkeefe@cooley.com

amace@cooley.com

FOR PATENT OWNER:

Amy Simpson

Robert Jain

HOLLAND & KNIGHT LLP

amy.simpson@hklaw.com

robert.jain@hklaw.com

Director_PTABDecision_Review@uspto.gov
571.272.7822

Paper 13
Date: January 12, 2026

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE OFFICE OF THE UNDER SECRETARY OF COMMERCE
FOR INTELLECTUAL PROPERTY AND DIRECTOR OF THE
UNITED STATES PATENT AND TRADEMARK OFFICE

GOOGLE LLC,
Petitioner,

v.

CELLULAR SOUTH, INC.,
Patent Owner.

IPR2025-00875 (Patent 9,940,972 B2)
IPR2025-00876 (Patent 10,218,954 B2)¹

Before JOHN A. SQUIRES, *Under Secretary of Commerce for Intellectual
Property and Director of the United States Patent and Trademark Office.*

ORDER

¹ This order applies to each of the above-listed proceedings.

IPR2025-00875 (Patent 9,940,972 B2)
IPR2025-00876 (Patent 10,218,954 B2)

The Office received a request for Director Review of the Decision denying institution in each of the above-captioned cases and an authorized response to each request. *See* Papers 11, 12.²

Having considered the requests and responses, it is:

ORDERED that the requests for Director Review are denied.

² Citations are to the record in IPR2025-00875. The parties filed similar papers in IPR2025-00876.

IPR2025-00875 (Patent 9,940,972 B2)
IPR2025-00876 (Patent 10,218,954 B2)

FOR PETITIONER:

Heidi L. Keefe
Andrew C. Mace
COOLEY LLP
hkeefe@cooley.com
amace@cooley.com

FOR PATENT OWNER:

Amy E. Simpson
Robert K. Jain
HOLLAND & KNIGHT LLP
amy.simpson@hklaw.com
robert.jain@hklaw.com

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE OFFICE OF THE UNDER SECRETARY OF COMMERCE
FOR INTELLECTUAL PROPERTY AND DIRECTOR OF THE
UNITED STATES PATENT AND TRADEMARK OFFICE

GOOGLE LLC,
Petitioner,

v.

CELLULAR SOUTH, INC.,
Patent Owner.

IPR2025-00875 (Patent 9,940,972 B2)
IPR2025-00876 (Patent 10,218,954 B2)
IPR2025-00877 (Patent 11,126,853 B2)

Before KALYAN K. DESHPANDE,¹ *Acting Chief Administrative Patent Judge.*

DECISION

Denying Institution of *Inter Partes* Review in IPR2025-00875 and
IPR2025-00876 and Referring the Petition to the Board in IPR2025-00877

¹ Coke Morgan Stewart, Deputy Under Secretary of Commerce for Intellectual Property and Deputy Director of the United States Patent and Trademark Office, is recused and took no part in this decision. See <https://www.uspto.gov/sites/default/files/documents/dsco-delegation.pdf>.

IPR2025-00875 (Patent 9,940,972 B2)
IPR2025-00876 (Patent 10,218,954 B2)
IPR2025-00877 (Patent 11,126,853 B2)

Cellular South, Inc. (“Patent Owner”) filed a request for discretionary denial (Paper 6, “DD Req.”) in the above-captioned cases, and Google LLC (“Petitioner”) filed an opposition (Paper 7, “DD Opp.”).²

After considering the parties’ arguments and the record, and in view of all relevant considerations, discretionary denial of institution is appropriate in IPR2025-00875 and IPR2025-00876, but is not appropriate in IPR2025-00877. This determination is based on the totality of the evidence and arguments the parties have presented.

Some factors counsel against discretionary denial. For example, the projected final written decision due date is December 16, 2026. DD Req. 9–10. The district court trial is scheduled to begin March 15, 2027. *Id.* at 6; DD Opp. 4. As such, it is likely that a final written decision in this proceeding will issue before the district court trial occurs, reducing the risk of duplication of efforts and inconsistent outcomes. Additionally, the patent challenged in IPR2025-00877 has not been in force for a significant amount of time (issued in 2021). Early challenges favor robust, predictable patent rights and weigh against discretionary denial.

The patents challenged in IPR2025-00875 and IPR2025-00876 present different circumstances. In particular, these challenged patents have been in force for seven and six years, respectively, creating settled expectations for Patent Owner, and Petitioner does not provide persuasive reasoning why an *inter partes* review is an appropriate use of Board resources. *Dabico Airport Sols. Inc. v. AXA Power ApS*, IPR2025-00408,

² Citations are to papers in IPR2025-00875. The parties filed similar papers in IPR2025-00876 and IPR2025-00877.

IPR2025-00875 (Patent 9,940,972 B2)
IPR2025-00876 (Patent 10,218,954 B2)
IPR2025-00877 (Patent 11,126,853 B2)

Paper 21 at 2–3 (Director June 18, 2025). In the absence of any such information, the Office is disinclined to disturb the settled expectations of Patent Owner.

Although certain arguments are highlighted above, the determinations in this Decision are based on a holistic assessment of all of the evidence and arguments presented. Accordingly, the Petitions in IPR2025-00875 and IPR2025-00876 are denied under 35 U.S.C. § 314(a), and the Petition in IPR2025-00877 is referred to the Board to handle the case in the normal course, including by issuing a decision on institution addressing the merits and other non-discretionary considerations, as appropriate.

In consideration of the foregoing, it is:

ORDERED that Patent Owner’s requests for discretionary denial in IPR2025-00875 and IPR2025-00876 are *granted*;

FURTHER ORDERED that the Petitions in IPR2025-00875 and IPR2025-00876 are *denied*, and no trial is instituted;

FURTHER ORDERED that Patent Owner’s request for discretionary denial in IPR2025-00877 is *denied*;

FURTHER ORDERED that the Petition in IPR2025-00877 is referred to the Board; and

FURTHER ORDERED that neither party shall file a request for rehearing or Director Review of the decision to deny Patent Owner’s request for discretionary denial in IPR2025-00877 until the Board issues a decision on institution.

IPR2025-00875 (Patent 9,940,972 B2)

IPR2025-00876 (Patent 10,218,954 B2)

IPR2025-00877 (Patent 11,126,853 B2)

FOR PETITIONER:

Heidi Keefe

Andrew Mace

COOLEY LLP

hkeefe@cooley.com

amace@cooley.com

FOR PATENT OWNER:

Amy Simpson

Robert Jain

HOLLAND & KNIGHT LLP

amy.simpson@hklaw.com

robert.jain@hklaw.com

Director_PTABDecision_Review@uspto.gov
571.272.7822

Paper 13
Date: January 12, 2026

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE OFFICE OF THE UNDER SECRETARY OF COMMERCE
FOR INTELLECTUAL PROPERTY AND DIRECTOR OF THE
UNITED STATES PATENT AND TRADEMARK OFFICE

GOOGLE LLC,
Petitioner,

v.

CELLULAR SOUTH, INC.,
Patent Owner.

IPR2025-00875 (Patent 9,940,972 B2)
IPR2025-00876 (Patent 10,218,954 B2)¹

Before JOHN A. SQUIRES, *Under Secretary of Commerce for Intellectual
Property and Director of the United States Patent and Trademark Office.*

ORDER

¹ This order applies to each of the above-listed proceedings.

IPR2025-00875 (Patent 9,940,972 B2)
IPR2025-00876 (Patent 10,218,954 B2)

The Office received a request for Director Review of the Decision denying institution in each of the above-captioned cases and an authorized response to each request. *See* Papers 11, 12.²

Having considered the requests and responses, it is:

ORDERED that the requests for Director Review are denied.

² Citations are to the record in IPR2025-00875. The parties filed similar papers in IPR2025-00876.

IPR2025-00875 (Patent 9,940,972 B2)
IPR2025-00876 (Patent 10,218,954 B2)

FOR PETITIONER:

Heidi L. Keefe
Andrew C. Mace
COOLEY LLP
hkeefe@cooley.com
amace@cooley.com

FOR PATENT OWNER:

Amy E. Simpson
Robert K. Jain
HOLLAND & KNIGHT LLP
amy.simpson@hklaw.com
robert.jain@hklaw.com

(12) **United States Patent**
Lakhani et al.

(10) **Patent No.:** **US 9,940,972 B2**
(45) **Date of Patent:** **Apr. 10, 2018**

(54) **VIDEO TO DATA**

(71) Applicant: **Cellular South, Inc.**, Ridgeland, MS (US)

(72) Inventors: **Naeem Lakhani**, Croyden (GB);
Bartlett Wade Smith, IV, Madison, MS (US)

(73) Assignee: **Cellular South, Inc.**, Ridgeland, MS (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 289 days.

(21) Appl. No.: **14/175,741**

(22) Filed: **Feb. 7, 2014**

(65) **Prior Publication Data**

US 2015/0050010 A1 Feb. 19, 2015

Related U.S. Application Data

(60) Provisional application No. 61/866,175, filed on Aug. 15, 2013.

(51) **Int. Cl.**

G10L 15/00 (2013.01)
G10L 25/00 (2013.01)
G06F 17/30 (2006.01)
G06T 7/00 (2017.01)
G11B 27/036 (2006.01)
G10L 25/57 (2013.01)
H04N 21/4402 (2011.01)
G10L 15/26 (2006.01)
G06K 9/00 (2006.01)

(52) **U.S. Cl.**

CPC **G11B 27/036** (2013.01); **G10L 25/57** (2013.01); **H04N 21/440236** (2013.01); **G06F 17/30** (2013.01); **G06K 9/00456** (2013.01); **G06K 2209/01** (2013.01); **G10L 15/26** (2013.01)

(58) **Field of Classification Search**

CPC G10L 15/18; G10L 15/26; G10L 15/265; G10L 25/48; G10L 25/51; G10L 25/54; G10L 25/57; G06F 17/30; G06F 17/30058; G06F 17/3074; G06F 17/30781; G06T 7/0081; G06K 9/00456; G06K 2209/01
USPC 704/231, 270, 270.1, 275; 382/100, 173, 382/176, 177; 345/418; 348/61
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,019,163 B2 * 9/2011 Momosaki G06F 17/30743 369/1
8,682,739 B1 * 3/2014 Feinstein G06Q 30/06 705/26.1
2003/0091237 A1 * 5/2003 Cohen-Solal G06K 9/3266 382/204
2003/0112261 A1 * 6/2003 Zhang 345/716
2003/0112265 A1 * 6/2003 Zhang 345/723
2003/0218696 A1 * 11/2003 Bagga G06F 17/30796 348/700
2005/0108004 A1 * 5/2005 Otani et al. 704/205
(Continued)

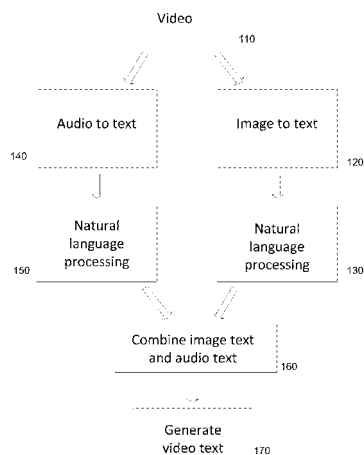
Primary Examiner — Paras D Shah

(74) *Attorney, Agent, or Firm* — Steptoe & Johnson LLP

(57) **ABSTRACT**

A method and system can generate video content from a video. The method and system can include generating audio files and image files from the video, distributing the audio files and the image files across a plurality of processors and processing the audio files and the image files in parallel. The audio files associated with the video to text and the image files associated with the video to video content can be converted. The text and the video content can be cross-referenced with the video.

20 Claims, 5 Drawing Sheets



US 9,940,972 B2

Page 2

(56)

References Cited

U.S. PATENT DOCUMENTS

2006/0179453	A1 *	8/2006	Kadie et al.	725/34
2006/0212897	A1 *	9/2006	Li et al.	725/32
2007/0112630	A1 *	5/2007	Lau et al.	705/14
2008/0120646	A1 *	5/2008	Stern et al.	725/34
2008/0276266	A1 *	11/2008	Huchital et al.	725/32
2009/0006191	A1 *	1/2009	Arankalle et al.	705/14
2009/0006375	A1 *	1/2009	Lax et al.	707/5
2009/0199235	A1 *	8/2009	Surendran et al.	725/34
2011/0292992	A1 *	12/2011	Srivara	375/240.01
2012/0042052	A1 *	2/2012	Ma	G06T 1/00 709/219
2012/0254917	A1 *	10/2012	Burkitt et al.	725/40
2013/0346144	A1 *	12/2013	Ferren	G06K 9/3266 705/7.29
2014/0314391	A1 *	10/2014	Kim	G11B 27/11 386/248
2015/0019206	A1 *	1/2015	Wilder	G06K 9/00302 704/9

* cited by examiner

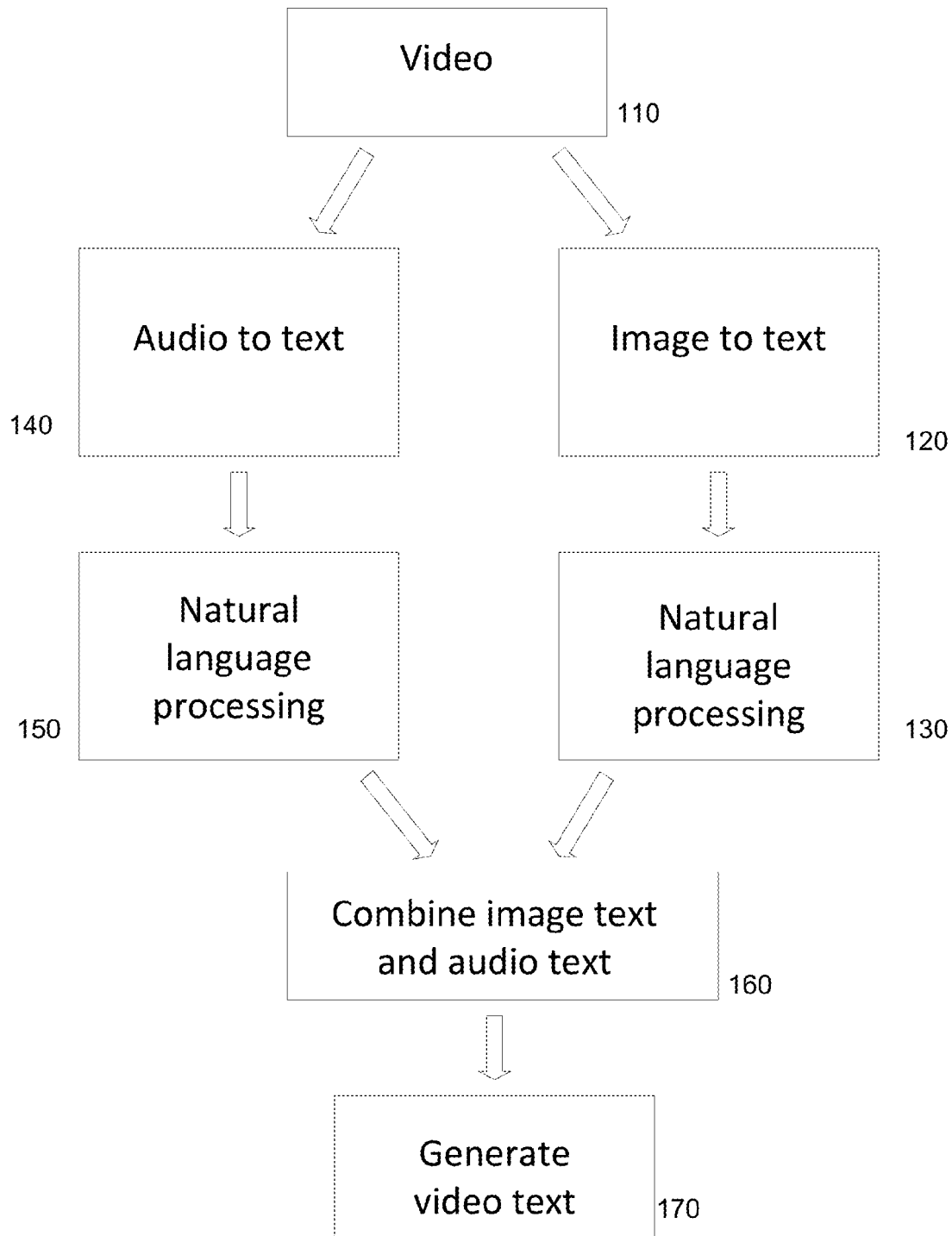


Figure 1

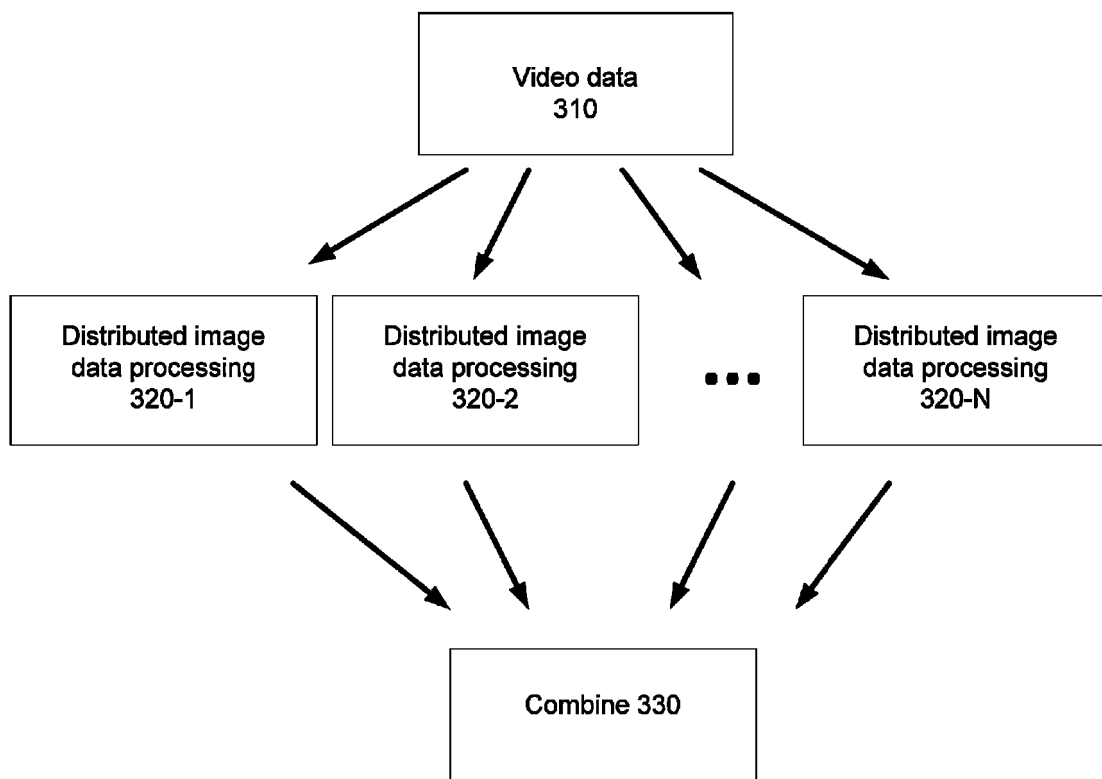


Figure 2

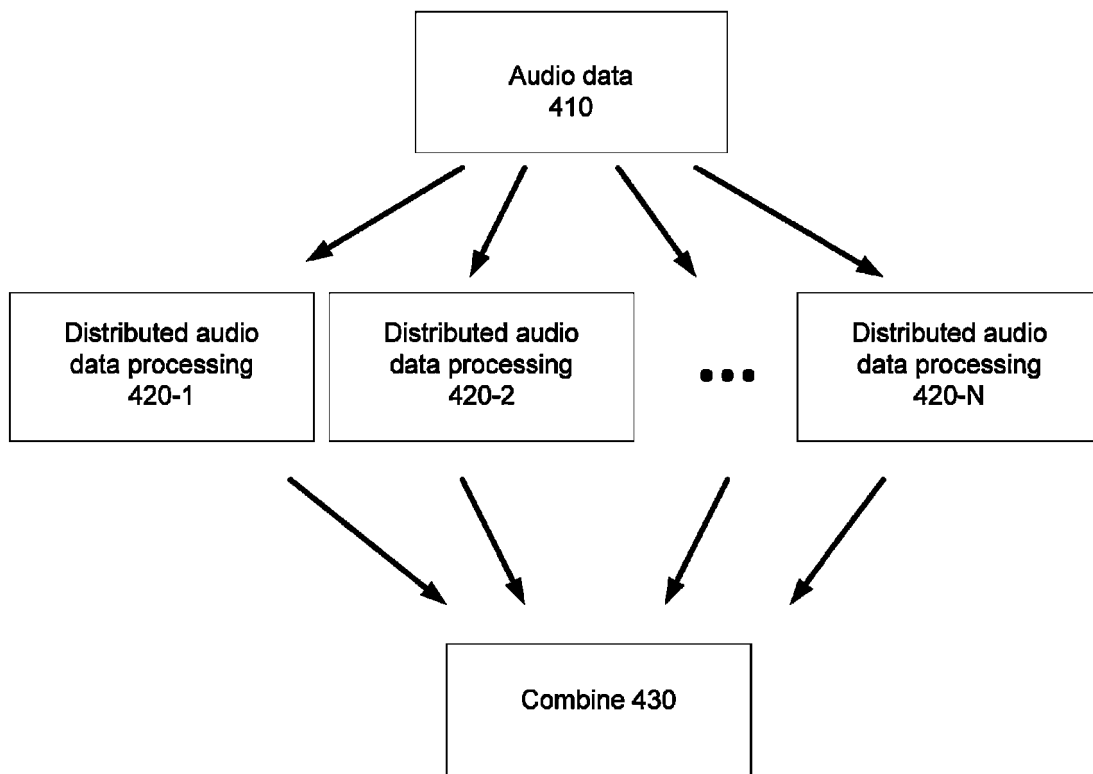


Figure 3

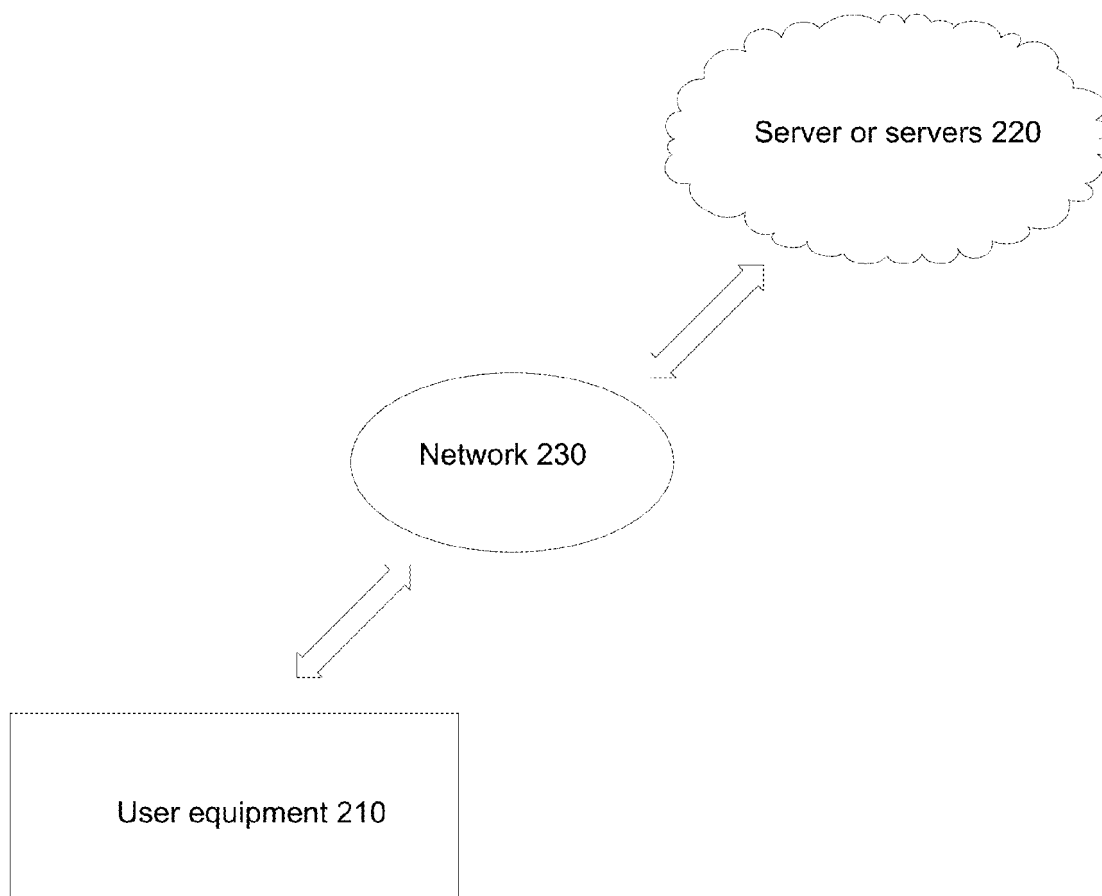


Figure 4

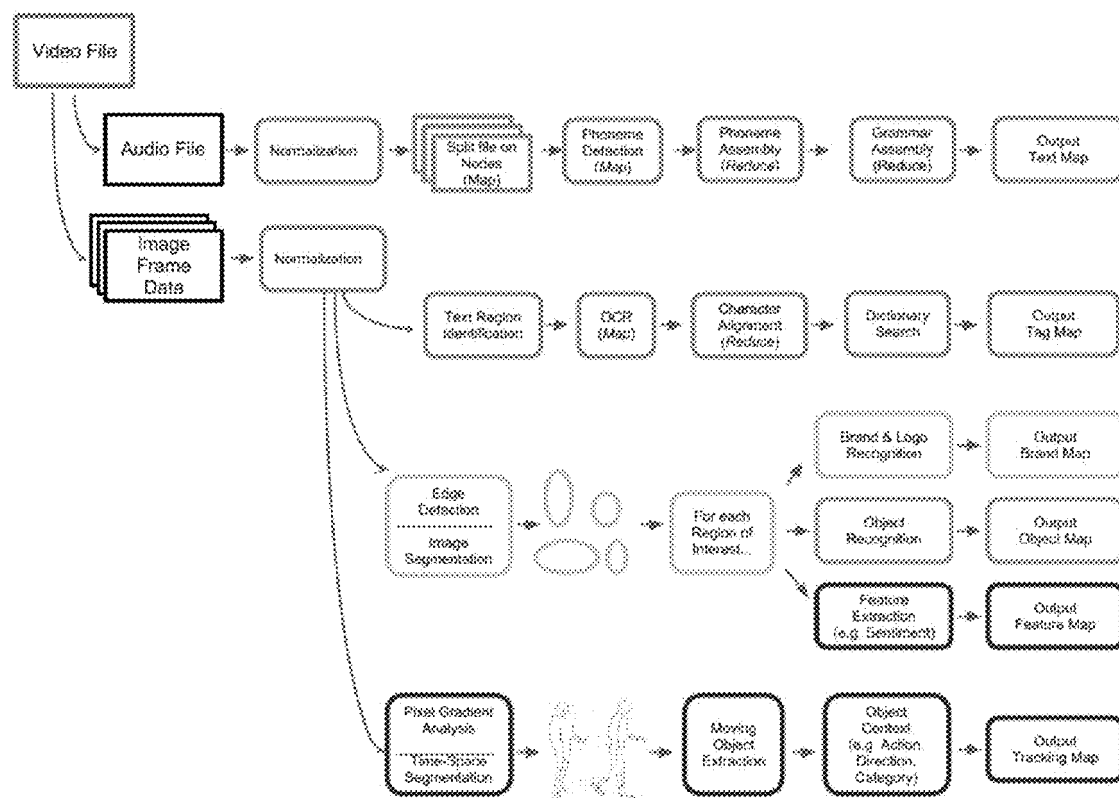


Figure 5

US 9,940,972 B2

1

VIDEO TO DATA

CLAIM OF PRIORITY

This application claims priority to U.S. Provisional Patent Application No. 61/866,175, filed on Aug. 15, 2013, which is incorporated by reference in its entirety.

TECHNICAL FIELD

The present invention relates to a method and a system for generating various and useful data from videos.

BACKGROUND

In the field of image contextualization, distributed reverse image similarity searching can be used to identify images similar to a target image. Reverse image searching can find exactly matching images as well as flipped, cropped, and altered versions of the target image. Distributed reverse image similarity searching can be used to identify symbolic similarity within images. Audio-to-text algorithms can be used to transcribe text from audio. An exemplary application is note-taking software. Audio-to-text, however, lacks semantic and contextual language understanding.

SUMMARY

The present invention is generally directed to a method to generate data from video content, such as text and/or image-related information. A server executing the method can be directed by a program stored on a non-transitory computer-readable medium. The video text can be, for example, a context description of the video.

An aspect of the method can include generating text from an image of the video, converting audio associated with the video to text, extracting topics from the text converted from the audio, cross-referencing the text generated from the image of the video and the topics extracted from audio associated with the video, and generating video text based on a result of the cross-referencing.

In some embodiments, natural language processing can be applied to the generation of text from an image of the video, converting audio associated with the video to text, or both.

In other embodiments, the text from the image of the video can be generated by identifying context, a symbol, a brand, a feature, an object, and/or a topic in the image of the video.

In yet other embodiments, the text from the image can be generated by first segmenting images of the video, and then converting the segments of images to text in parallel. The text from the audio can be generated by first segmenting images of the audio, and then converting the segments of images to text in parallel. The audio can be segmented at spectrum thresholds.

In some embodiments, the method can include generating advertising based on video content. The video content can be text, context, symbols, brands, features, objects, and/or topics related to or found in the video. An advertisement can be placed at a specific time in the video based on the video content and/or section symbol of a video image. In some embodiments, the method can include directing when one or more advertisements can be placed in a predetermined context at a preferred time.

DESCRIPTION OF THE DRAWINGS

The present invention is further described in the detailed description which follows, in reference to the noted plurality

2

of drawings by way of non-limiting examples of certain embodiments of the present invention, in which like numerals represent like elements throughout the several views of the drawings, and wherein:

FIG. 1 illustrates an embodiment of present invention.

FIG. 2 illustrates an embodiment of image data processing.

FIG. 3 illustrates an embodiment of audio data processing.

FIG. 4 illustrates another embodiment of present invention.

FIG. 5 illustrates various exemplary embodiments of present invention.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

A detailed explanation of the system and method according to exemplary embodiments of the present invention are described below. Exemplary embodiments described, shown, and/or disclosed herein are not intended to limit the claims, but rather, are intended to instruct one of ordinary skill in the art as to various aspects of the invention. Other embodiments can be practiced and/or implemented without departing from the scope and spirit of the claimed invention.

The present invention is generally directed to system, device, and method of generating content from video files, such as text and information relating to context, symbols, brands, features, objects, faces and/or topics found in the images of such videos. In an embodiment, the video-to-content engine can perform the functions directed by programs stored in a computer-readable medium. That is, the embodiments may take the form of a hardware embodiment (including circuits), a software embodiment, or an embodiment combining software and hardware. The present invention can take the form of a computer-program product that includes computer-useable instructions embodied on one or more computer-readable media.

The various video-to-content techniques, methods, and systems described herein can be implemented in part or in whole using computer-based systems and methods. Additionally, computer-based systems and methods can be used to augment or enhance the functionality described herein, increase the speed at which the functions can be performed, and provide additional features and aspects as a part of or in addition to those described elsewhere in this document. Various computer-based systems, methods and implementations in accordance with the described technology are presented below.

A video-to-content engine can be embodied by the a general-purpose computer or a server and can have an internal or external memory for storing data and programs such as an operating system (e.g., DOS, Windows 2000™, Windows XP™, Windows NT™, OS/2, UNIX or Linux) and one or more application programs. Examples of application programs include computer programs implementing the techniques described herein for lyric and multimedia customization, authoring applications (e.g., word processing programs, database programs, spreadsheet programs, or graphics programs) capable of generating documents or other electronic content; client applications (e.g., an Internet Service Provider (ISP) client, an e-mail client, or an instant messaging (IM) client) capable of communicating with other computer users, accessing various computer resources, and viewing, creating, or otherwise manipulating electronic content; and browser applications (e.g., Microsoft's Internet Explorer) capable of rendering standard Internet content and

US 9,940,972 B2

3

other content formatted according to standard protocols such as the Hypertext Transfer Protocol (HTTP). One or more of the application programs can be installed on the internal or external storage of the general-purpose computer. Alternatively, application programs can be externally stored in or performed by one or more device(s) external to the general-purpose computer.

The general-purpose computer or server may include a central processing unit (CPU) for executing instructions in response to commands, and a communication device for sending and receiving data. One example of the communication device can be a modem. Other examples include a transceiver, a communication card, a satellite dish, an antenna, a network adapter, or some other mechanism capable of transmitting and receiving data over a communications link through a wired or wireless data pathway.

The general-purpose computer or server may also include an input/output interface that enables wired or wireless connection to various peripheral devices. In one implementation, a processor-based system of the general-purpose computer can include a main memory, preferably random access memory (RAM), and can also include a secondary memory, which may be a tangible computer-readable medium. The tangible computer-readable medium memory can include, for example, a hard disk drive or a removable storage drive, a flash based storage system or solid-state drive, a floppy disk drive, a magnetic tape drive, an optical disk drive (Blu-Ray, DVD, CD drive), magnetic tape, paper tape, punched cards, standalone RAM disks, Iomega Zip drive, etc. The removable storage drive can read from or write to a removable storage medium. A removable storage medium can include a floppy disk, magnetic tape, optical disk (Blu-Ray disc, DVD, CD) a memory card (Compact-Flash card, Secure Digital card, Memory Stick), paper data storage (punched card, punched tape), etc., which can be removed from the storage drive used to perform read and write operations. As will be appreciated, the removable storage medium can include computer software or data.

In alternative embodiments, the tangible computer-readable medium memory can include other similar means for allowing computer programs or other instructions to be loaded into a computer system. Such means can include, for example, a removable storage unit and an interface. Examples of such can include a program cartridge and cartridge interface (such as the found in video game devices), a removable memory chip (such as an EPROM or flash memory) and associated socket, and other removable storage units and interfaces, which allow software and data to be transferred from the removable storage unit to the computer system.

An embodiment of video-to-text engine operation is illustrated in FIG. 1. At 110, a video stream is presented. The video stream may be in format of (but not limited to): Advanced Video Codec High Definition (AVCHD), Audio Video Interlaced (AVI), Flash Video Format (FLU), Motion Picture Experts Group (MPEG), Windows Media Video (WMV), or Apple QuickTime (MOV), h.264 (MP4).

The engine can extract audio data and image data (e.g. images or frames forming the video) from the video stream.

In some embodiments, the video stream and the extracted image data can be stored in a memory or storage device such as those discussed above. A copy of the extracted image data can be used for processing.

At 120, the video-to-text engine performs an image data processing on the video stream. An example of the image data processing is illustrated in FIG. 2. In FIG. 2, the image data 310 can be segmented into N segments and processed

4

in parallel (e.g., distributed processing 320-1 to 320-N), allowing for near real-time processing.

An example of the video image data processing can be symbol (or object) based. Using image processing technique such as color edge detection, a symbol of a screen or an image of the video can be isolated. The symbol can be identified using an object template database. For example, the symbol includes 4 legs and a tail, and when matched with the object template database, the symbol may be identified as a dog. The object template database can be adaptive and therefore, the performance would improve with usage.

Other image data processing techniques may include image extraction, high-level vision and symbol detection, figure-ground separation, depth and motion perception.

Another example of video image processing can be color segmentation. The colors of an image (e.g., a screen) of the video can be segmented or grouped. The result can be compared to a database using color similarity matching.

Based on the identified symbol, a plurality of instances of the symbol can be compared to a topic database to identify a topic (such as an event). For example, the result may identify the dog (symbol) as running or jumping. The topic database can be adaptive to improve its performance with usage.

Thus, using the processing example above, text describing a symbol of the video and topic relating to the symbol may be generated. Data generated from an image and/or from audio transcription can be time stamped, for example, according to when it appeared, was heard, and/or according to the video frame from which it was pulled.

At 330, the engine combines the topics as an array of keys and values with respect to the segments. The engine can segment the topics over a period of time and weight the strength of each topic. Further, the engine applies the topical meta-data to the original full video. The image topics can be stored as topics for the entire video or each image segment. The topic generation process can be repeated for all identifiable symbols in a video in a distributed process. The outcome would be several topical descriptors of the content within a video. An example of the aggregate information that would be derived using the above example would be understanding that the video presented a dog, which was jumping, on the beach, with people, by a resort.

Natural language processing can be useful in creating an intuitive and/or user-friendly computer-human interaction. In some embodiments, the system can select semantics or topics, following certain rules, from a plurality of possible semantics or topics, can give them weight based on strength of context, and/or can do this in a distributed environment. The natural language processing can be augmented and/or improved by implementing machine-learning. A large training set of data can be obtained from proprietary or publicly available resources. For example, CBS News maintains a database of segments and episodes of "60-Minutes" with full transcripts, which can be useful for building a training set and for unattended verification of audio segmentation. The machine learning can include ensemble learning based on the concatenation of several classifiers, i.e. cascade classifiers.

At 130, an optional step of natural language processing can be applied to the image text. For example, based on dictionary, grammar, and a knowledge database, the text extracted from video images can be modified as the video-to-text engine selects primary semantics from a plurality of possible semantics. In some embodiments, the system and method can incorporate a Fourier transform of the audio

US 9,940,972 B2

5

signal. Such filtering can improve silence recognition, which can be useful for determining proper placement of commas and periods in the text file.

In parallel, at **140**, the video-to-text engine can perform audio-to-text processing on audio data associated with the video. For example, for a movie video, the associated audio may be the dialog or even background music.

An example of the audio data processing is illustrated in FIG. 3. In FIG. 3, the audio data **410** can be segmented into N segments and processed in parallel (e.g., distributed processing **420-1** to **420-N**), allowing for near real-time processing.

In some embodiments, the segmentation can be performed by a fixed period of time. In another example, quiet periods in the audio data can be detected, and the segmentation can be defined by the quiet periods. For example, the audio data can be processed and converted into a spectrum. Locations where the spectrum volatility is below a threshold can be detected and segmented. Such locations can represent silence or low audio activities in the audio data. The quiet periods in the audio data can be ignored, and the processing requirements thereof can be reduced.

Audio data and/or segments of audio data can be stored in, for example, memory or storage device discussed above. Copies of the audio segments can be sent to audio processing.

The audio data for each segment can be translated into text in parallel, for example through distributed computing, which can reduce processing time. Various audio analysis tools and processes can be used, such as audio feature detection and extraction, audio indexing, hashing and searching, semantic analysis, and synthesis.

At **430**, text for a plurality of segments can then be combined. The combination can result in segmented transcripts and/or a full transcripts of the audio data. In an embodiment, the topics in each segment can be extracted. When combined, the topics in each segment can be given a different weight.

The audio topics can be stored as topics for the entire video or each audio segment.

At **150**, an optional step of natural language processing can be applied to the text. For example, based on dictionary, grammar, and/or a knowledge database, the text extract from the audio stream of a video can be given context, an applied sentiment, and topical weightings.

At **160**, the topics generated from an image or a frame and the topics extracted from audio can be combined. The text can be cross-referenced, and topics common to both texts would be given additional weights. At **170**, the video-to-text engine generates video text, such as text describing the content of the video, using the result of the combined texts and cross reference. For example, key words indicating topic and semantic that appear in both texts can be selected or emphasized.

FIG. 4 illustrates another embodiment of the present invention. User equipment (UE) **210** can communicate with a server or servers **220** via a network **230**. An exemplary embodiment of the system can be implemented over a cloud computing network.

The UE **210** can include, for example, a laptop, a tablet, a mobile phone, a personal digital assistant (PDA), a keyboard, a display monitor with or without a touch screen input, and an audiovisual input device. In another implementation, the peripheral devices may themselves include the functionality of the general-purpose computer. For example, the mobile phone or the PDA may include computing and networking capabilities and function as a general

6

purpose computer by accessing a network and communicating with other computer systems.

The server **220** includes the general purpose computer discussed above.

The network **230** includes, for example, the Internet, the World Wide Web, WANs, LANs, analog or digital wired and wireless telephone networks (e.g., Public Switched Telephone Network (PSTN), Integrated Services Digital Network (ISDN), and Digital Subscriber Line (xDSL)), radio, television, cable, or satellite systems, and other delivery mechanisms for carrying data. A communications link can include communication pathways that enable communications through one or more networks.

In some embodiments, a video-to-content engine can be embodied in a server or servers **220**. The UE **210**, for example, requests an application relating to the video stream. The servers **220** perform the audio-to-text process on the segmented audio in parallel. The distributed audio-to-text processing reduces the overall response time. This method allows real-time audio-to-text conversion.

The UE **210** communicates with the server **220** via the network **230** for video stream application. The video-to-content engine can generate the video text as illustrated in FIG. 1. The server **220** then generates advertisement (text, images, or animation) based on the video text. In some embodiments, the server adds the advertisement to a specific symbol, image, frame, or a specific time in the video stream. The specific symbol, image, frame, or the specific time in the video stream can be selected based on the video text.

The server **220** can add the audio text to the video stream in real time (i.e. real time close caption).

The server **220** can generate video recommendation based on a database of the video text. In some embodiments, the server **220** can search videos based on the video text (e.g., via a database of video text). In this fashion, video search can be optimized. Applications for the video search optimization may include search engine optimization (SEO), search engine marketing (SEM), censorship and removal materials of copyright violation.

The video streams can be videos viewed by a user, and the server **220** generates a preference profile for the user using the video data.

In an embodiment, as shown in FIG. 5 for example, a server node can fetch a video file. For example, a URL can be used to fetch the video file from an Internet such as YouTube, and from such URL the video can be scraped. The server can divide the video into chunks of smaller data files for processing on several nodes of a cluster in parallel. For example, the video file can be separated into audio files and image frame files. Each of the types of files can be normalized.

The normalized audio files can be split into constituent files for processing and reduction in parallel by various nodes. Various reduction processes can be performed on the constituent audio files such as phoneme detection and assembly as well as grammar assembly. An output of the audio processing steps can be an extracted text map.

The normalized image frame files can be processed in order to extract various data maps, such as a text map, a tag map, a brand, an object map, a feature map, and/or a tracking map. Such maps can be achieved through various extraction steps. For example, the normalized image frame files can be analyzed for text identification and/or by optical character recognition. The data can be improved through a dictionary verification step. Various maps can be created based on edge detection and/or image segmentation techniques. Such techniques can be improved by focusing on regions of interest,

US 9,940,972 B2

7

for example based on brands, logos, objects, and/or features of interest. Additionally, or alternatively, pixel gradients of the normalized image frame files can be analyzed and/or the files can be segmented by temporal and/or spatial components, and thus, for example, allow extraction of motion within the video images, which in turn can be used for tracking.

An advantage of present embodiments includes the ability to provide real-time or faster-than-real-time content output. This can be achieved through one or more components and/or steps. For example, a video file can be distributed across at least two layers for processing. The audio can be converted to text on at least one layer, and the images can be processed on at least one other layer. In some embodiments, natural language processing can abstract topics, sentiments, temporal topic-tagging, and can be used for further optimization and/or machine learning. The layers can include node clusters for parallel processing chunks of the video file into the preferred content. In some embodiments, the files can be maintained and processed in parallel at each step, and then combined into a single data file as one of the terminal processing steps.

Present embodiments have wide application. For example, video indexing, reverse image lookup, video groupings and graph searches, and video similarity indexing, as described herein, can be used for searching, for classification, and for recommendations regarding processed videos. Law enforcement and security industries can implement embodiments for object recognition and motion detection. Media, entertainment, and industrial entities can implement embodiments to monitor for trademark infringement, captioning, advertising and targeting, brand and product monitoring and data collection, and marketing analytics. These exemplary implementation are not intended to be limiting, merely exemplary.

Additionally, or alternatively, to actively fetching and scraping a video, the system and method can be automated as a push system and/or a web crawling system. For example, the server can monitor an RSS feed or online content of specific providers, such as YouTube, Vimeo, the growing myriad of video-content creating websites, or other online video providers. Monitoring of published videos can be tailored to search for extracted data relevant to specific requesters. For example, a purveyor of certain products can be apprised in real-time of new content relevant to the products. Such relevant content can include the context in which the products are found in the video, the appearance of competing products, verification of product placement, and other useful information.

All of the methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the apparatus and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope or the invention. In addition, from the foregoing it will be seen that this invention is one well adapted to attain all the ends and objects set forth above, together with other advantages. It will be understood that certain features and sub-combinations are of utility and may be employed without reference to other features and sub-combinations. This is contemplated and within the scope of the appended claims. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit and scope of the invention as defined by the appended claims.

8

The invention claimed is:

1. A method to generate video data from a video comprising:
 - generating audio files and image files from the video;
 - distributing the image files across a plurality of processors and processing the image files in parallel, wherein processing the image files comprises extracting one or more objects and identifying the one or more objects;
 - processing the audio files;
 - converting audio files associated with the video to text;
 - converting the image files associated with the video to video data;
 - generating a topical meta-data that describes content of the video by deriving semantic information from the identification of the one or more objects and semantic information from the audio files;
 - adding the topical meta-data to the video; and
 - cross-referencing the text and the video data based on the generated topical meta-data to determine topics;
 - generating video text based on the cross-referencing, wherein the video text describes content of the video;
 - generating a text, image, or animation based on the video text; and
 - placing the text, image, or animation in the video.
2. The method according to claim 1, further comprising: generating a content-rich video based on the video, the text, and the video data.
3. The method according to claim 1, further comprising: applying natural language processing to the text to determine context associated with the video.
4. The method according to claim 2, further comprising: applying natural language processing to the text to extract the topical meta-data.
5. The method according to claim 1, further comprising: processing the image files to extract additional text.
6. The method according to claim 5, wherein the additional text is generated by segmenting the image files before processing the image files in parallel.
7. The method according to claim 1, further comprising: determining a motion associated with the one or more objects.
8. The method according to claim 1, further comprising segmenting the audio files before processing the audio files in parallel.
9. The method according to claim 7, wherein the audio files and the image files are segmented at spectrum thresholds.
10. The method according to claim 1, further comprising: generating an advertisement based on the text and the video data.
11. The method according to claim 10, further comprising: placing the advertisement in the video at a preferred time.
12. The method according to claim 6 wherein the additional text includes information regarding context associated with the video.
13. The method according to claim 6, wherein the additional text relates to a symbol appearing in the video.
14. The method according to claim 13, wherein the symbol is a brand logo, and wherein the additional text includes information regarding placement and time of appearance of the brand logo.
15. The method according to claim 1, wherein the one or more objects are letters appearing in the video.
16. The method according to claim 6, wherein the additional text relates to faces appearing in the video.

US 9,940,972 B2

9

10

- 17.** A system for extracting data from a video, comprising:
a computer processor having a plurality of processors for
parallel processing; and
a non-transitory computer readable medium containing
instructions directing the system to execute the steps of: 5
converting audio associated with the video to text;
converting images associated with the video to video data;
generating the video data by segmenting image files of the
video before processing the image files in parallel;
identifying one or more objects in the image files; 10
generating data topics, from the text and the video data,
that describe content of the video by deriving semantic
information from the identification of the one or more
objects and semantic information from the audio;
adding the data topics to the video as meta-data; 15
cross-referencing the text, the video data, and the topics
with the video based on the generated data topics;
generating a text, image, or animation based on the data
topics; and
placing the text, image, or animation in the video. 20
- 18.** The system according to claim **17**, wherein converting
the audio comprises natural language processing.
- 19.** The system according to claim **17**, the computer
directs the audio to be converted by at least one node of a
cluster and the computer directs the images to be converted 25
by at least one other node of the cluster in parallel.
- 20.** The server according to claim **17**, wherein the audio
and the images are segmented at spectrum thresholds.

* * * * *



US010218954B2

(12) **United States Patent**
Lakhani et al.

(10) **Patent No.:** **US 10,218,954 B2**

(45) **Date of Patent:** ***Feb. 26, 2019**

(54) **VIDEO TO DATA**

(71) Applicant: **CELLULAR SOUTH, INC.**,
Ridgeland, MS (US)

(72) Inventors: **Nacem Lakhani**, Croydon (GB);
Bartlett Wade Smith, IV, Madison,
MS (US); **Allison A. Talley**, Ridgeland,
MS (US)

(73) Assignee: **CELLULAR SOUTH, INC.**,
Ridgeland, MS (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 0 days.

This patent is subject to a terminal dis-
claimer.

(21) Appl. No.: **14/910,698**

(22) PCT Filed: **Feb. 7, 2015**

(86) PCT No.: **PCT/US2015/014940**

§ 371 (c)(1),

(2) Date: **Feb. 6, 2016**

(87) PCT Pub. No.: **WO2015/120351**

PCT Pub. Date: **Aug. 13, 2015**

(65) **Prior Publication Data**

US 2016/0358632 A1 Dec. 8, 2016

Related U.S. Application Data

(63) Continuation of application No. 14/175,741, filed on
Feb. 7, 2014, now Pat. No. 9,940,972.

(Continued)

(51) **Int. Cl.**

G10L 15/00 (2013.01)

G10L 25/00 (2013.01)

(Continued)

(52) **U.S. Cl.**

CPC **H04N 9/8715** (2013.01); **H04N 9/80**
(2013.01); **H04N 9/802** (2013.01); **G10L**
15/26 (2013.01)

(58) **Field of Classification Search**

CPC ... G01L 5/18; G01L 5/26; G01L 5/265; G10L
25/48; G10L 25/51; G10L 25/54;
(Continued)

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,536,413 B1 * 5/2009 Mohan G06F 17/3071

9,615,136 B1 * 4/2017 Emery H04N 21/47202

(Continued)

OTHER PUBLICATIONS

Notification Concerning Transmittal of International Preliminary
Report on Patentability (Chapter I of the Patent Cooperation Treaty)
dated Aug. 18, 2016, issued in International Application No. PCT/
US2015/014940.

(Continued)

Primary Examiner — Paras D Shah

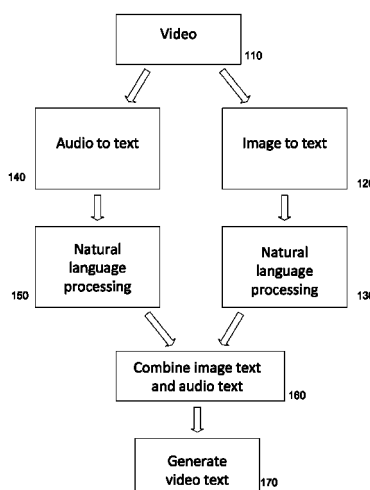
(74) *Attorney, Agent, or Firm* — Steptoe & Johnson LLP

(57)

ABSTRACT

A method and system can generate video content from a
video. The method and system can include generating audio
files and image files from the video, distributing the audio
files and the image files across a plurality of processors and
processing the audio files and the image files in parallel. The
audio files associated with the video to text and the image
files associated with the video to video content can be
converted. The text and the video content can be cross-
referenced with the video.

18 Claims, 9 Drawing Sheets



US 10,218,954 B2

Page 2

Related U.S. Application Data

- (60) Provisional application No. 62/021,666, filed on Jul. 7, 2014, provisional application No. 61/866,175, filed on Aug. 15, 2013.

(51) **Int. Cl.**

G06F 17/30 (2006.01)
G06T 7/00 (2017.01)
H04N 9/87 (2006.01)
H04N 9/802 (2006.01)
H04N 9/80 (2006.01)
G10L 15/26 (2006.01)

(58) **Field of Classification Search**

CPC ... G10L 25/57; G06F 17/30; G06F 17/30058;
 G06F 17/3074; G06F 17/30781; G06T
 7/0081; G06K 2209/01
 USPC 704/231, 270, 270.1, 275; 382/100, 173,
 382/176, 177; 345/418; 348/61
 See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

9,652,477 B2 * 5/2017 Balinsky G06F 17/3028
 2001/0020954 A1 9/2001 Hull et al.
 2003/0112261 A1 * 6/2003 Zhang G11B 27/22
 715/716
 2003/0112265 A1 * 6/2003 Zhang G06F 17/30787
 715/723
 2005/0108004 A1 * 5/2005 Otani G10L 25/78
 704/205
 2006/0179453 A1 * 8/2006 Kadie G06Q 30/02
 725/34
 2006/0212897 A1 * 9/2006 Li H04H 60/58
 725/32

2007/0106685 A1 * 5/2007 Houh G06F 17/30796
 2007/0112630 A1 * 5/2007 Lau G06K 9/72
 705/14.1
 2008/0120646 A1 * 5/2008 Stern G06Q 30/02
 725/34
 2008/0232696 A1 * 9/2008 Kasahara G06K 9/00664
 382/224
 2008/0276266 A1 * 11/2008 Huchital G06Q 30/02
 725/32
 2009/0006191 A1 * 1/2009 Arankalle G06Q 30/02
 705/14.71
 2009/0006375 A1 * 1/2009 Lax H04N 21/435
 2009/0157407 A1 6/2009 Yamabe et al.
 2009/0199235 A1 * 8/2009 Surendran G06Q 30/02
 725/34
 2011/0069936 A1 3/2011 Johnson et al.
 2011/0292992 A1 * 12/2011 Sirivara H04N 21/23418
 375/240.01
 2012/0254917 A1 * 10/2012 Burkitt G06F 17/30817
 725/40
 2013/0174045 A1 7/2013 Sarukkai et al.
 2014/0314391 A1 * 10/2014 Kim G11B 27/11
 386/248
 2015/0019206 A1 * 1/2015 Wilder G06K 9/00302
 704/9

OTHER PUBLICATIONS

International Search Report dated Jun. 3, 2015, issued in International Application No. PCT/US2015/014940.
 Written Opinion of the International Searching Authority dated Jun. 3, 2015, issued in International Application No. PCT/US2015/014940.
 Deng, Yining, and B.S. Manjunath, "Content-based search of video using color, texture, and motion." Image Processing, 1997, retrieved from: <http://vision.ece.ucsb.edu/publications/97ICIPCBS.pdf>.

* cited by examiner

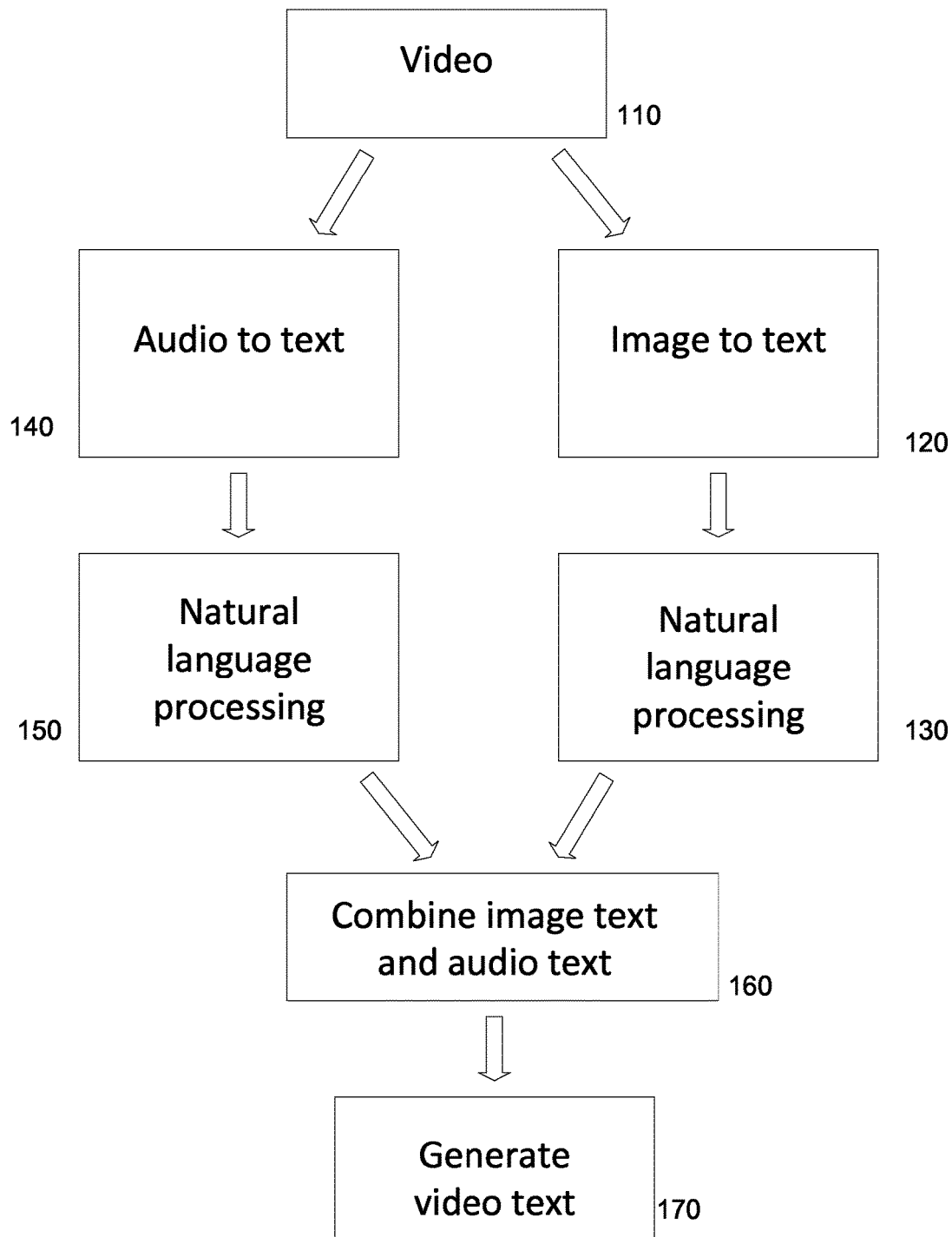


FIG. 1

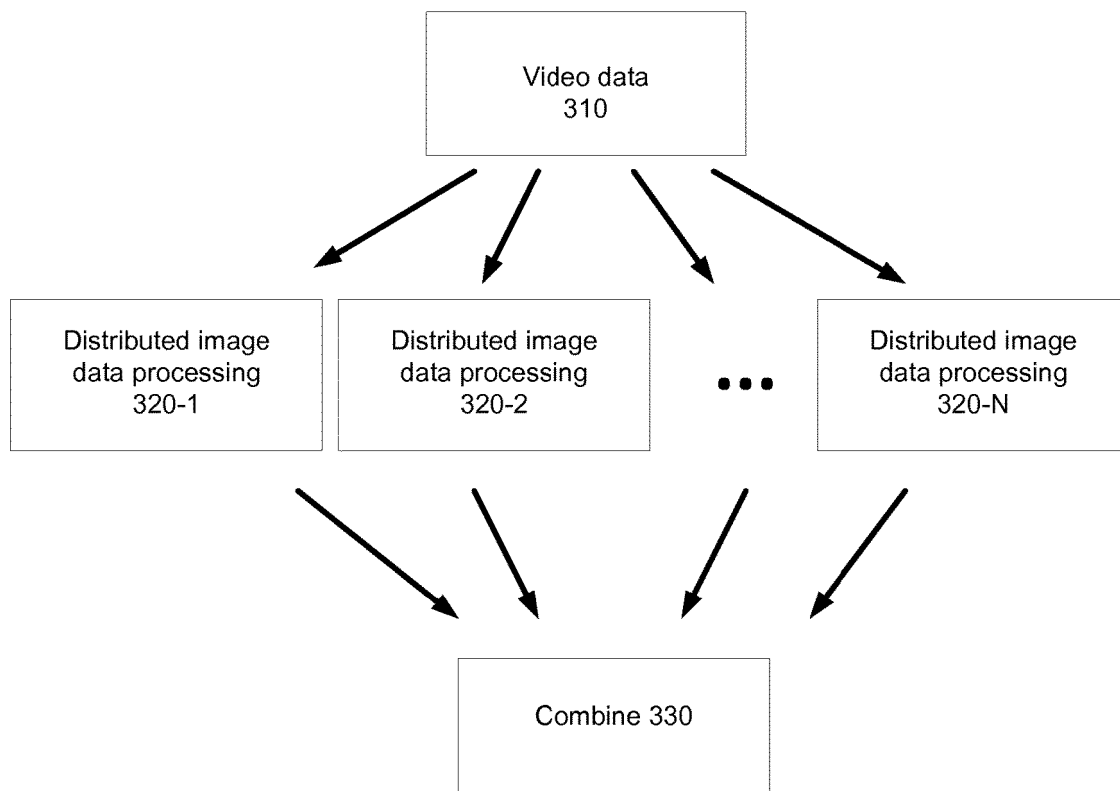


FIG. 2

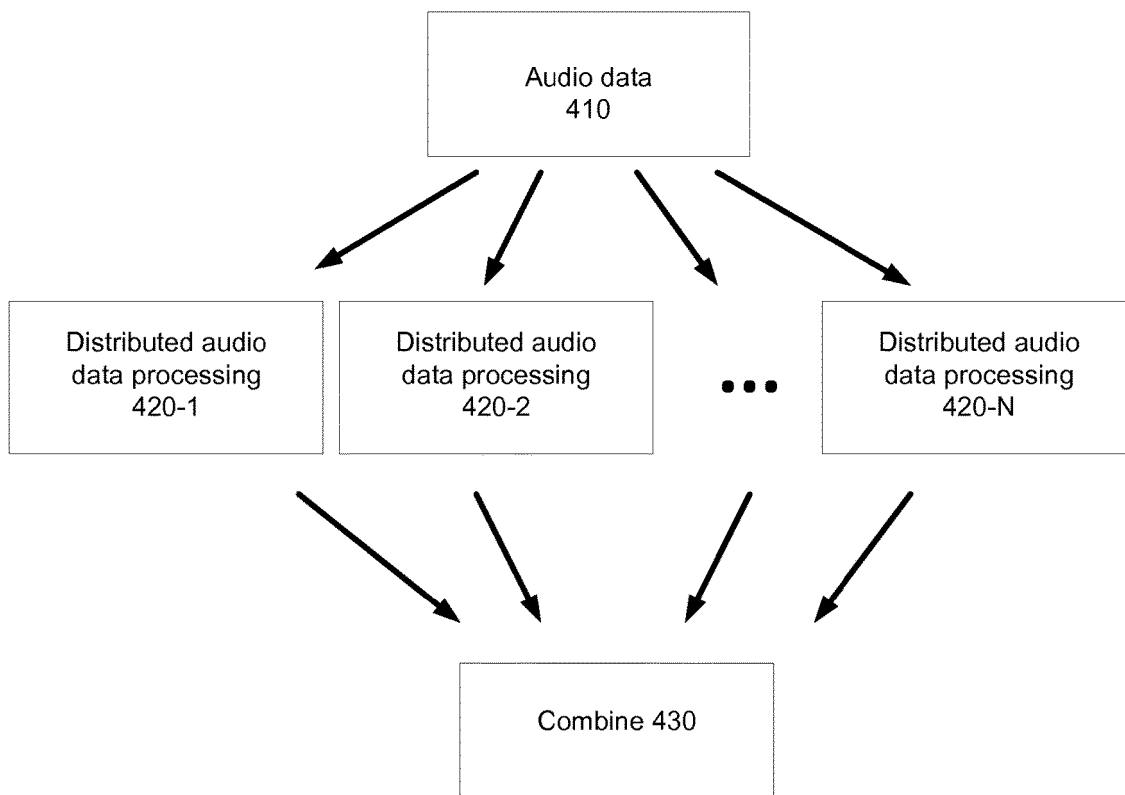


FIG. 3

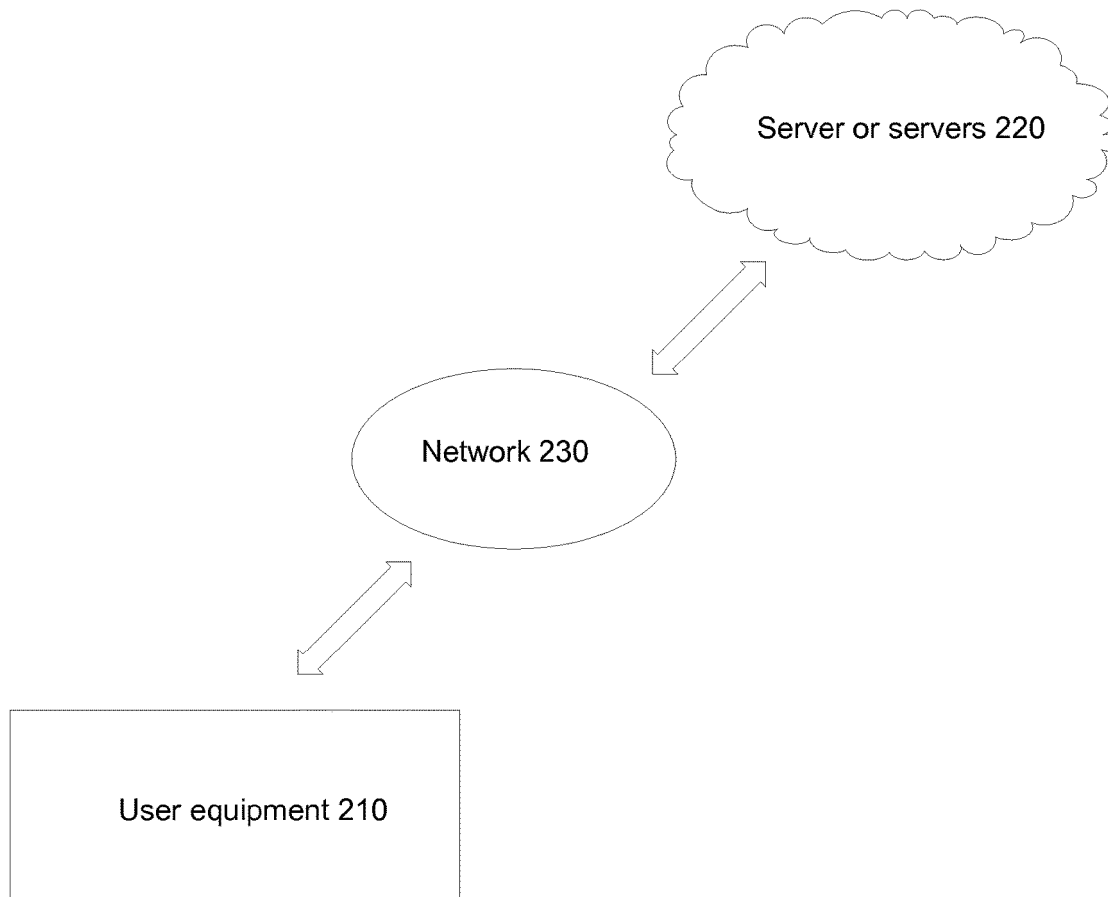


FIG. 4

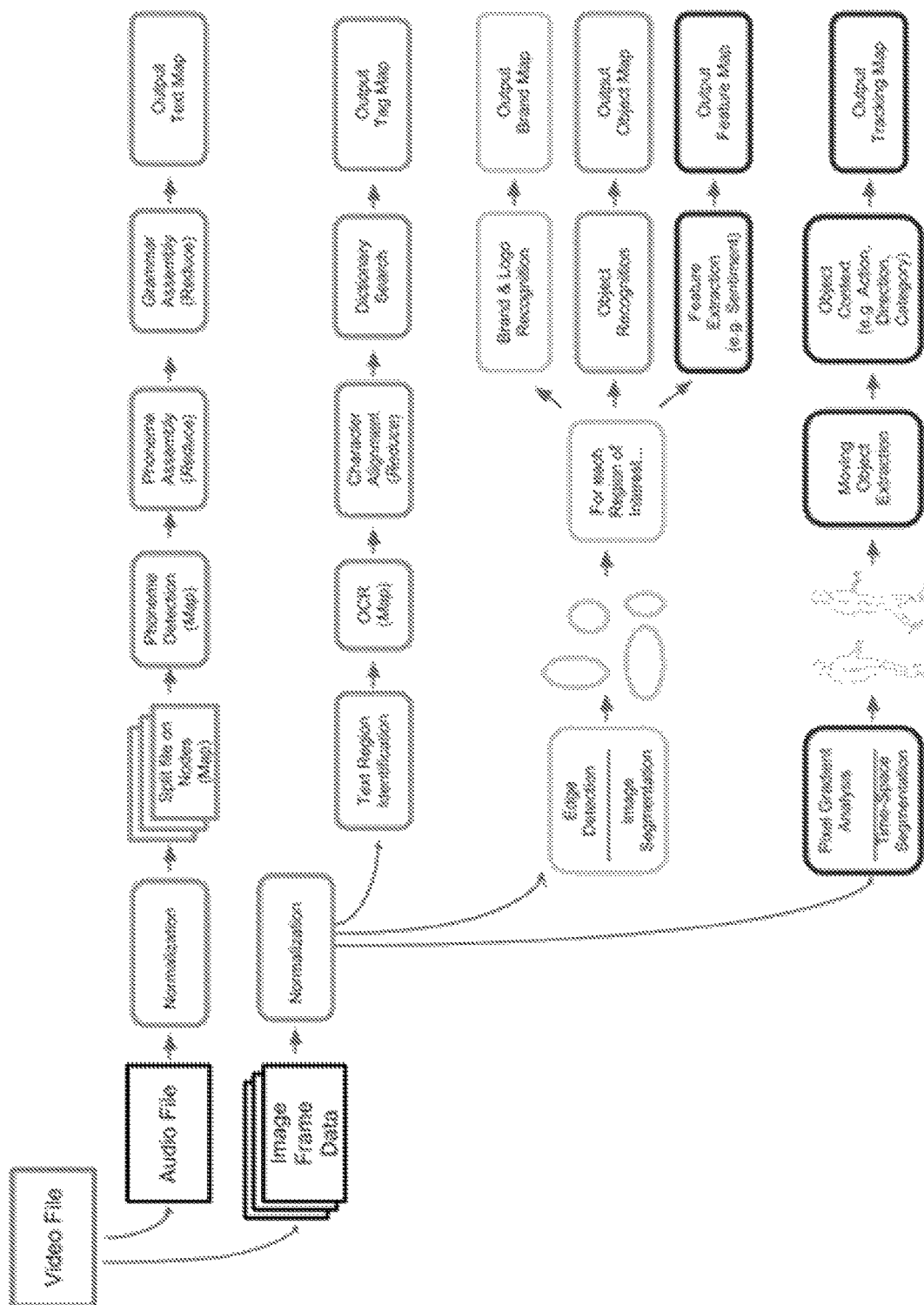


FIG. 5

Figure 6



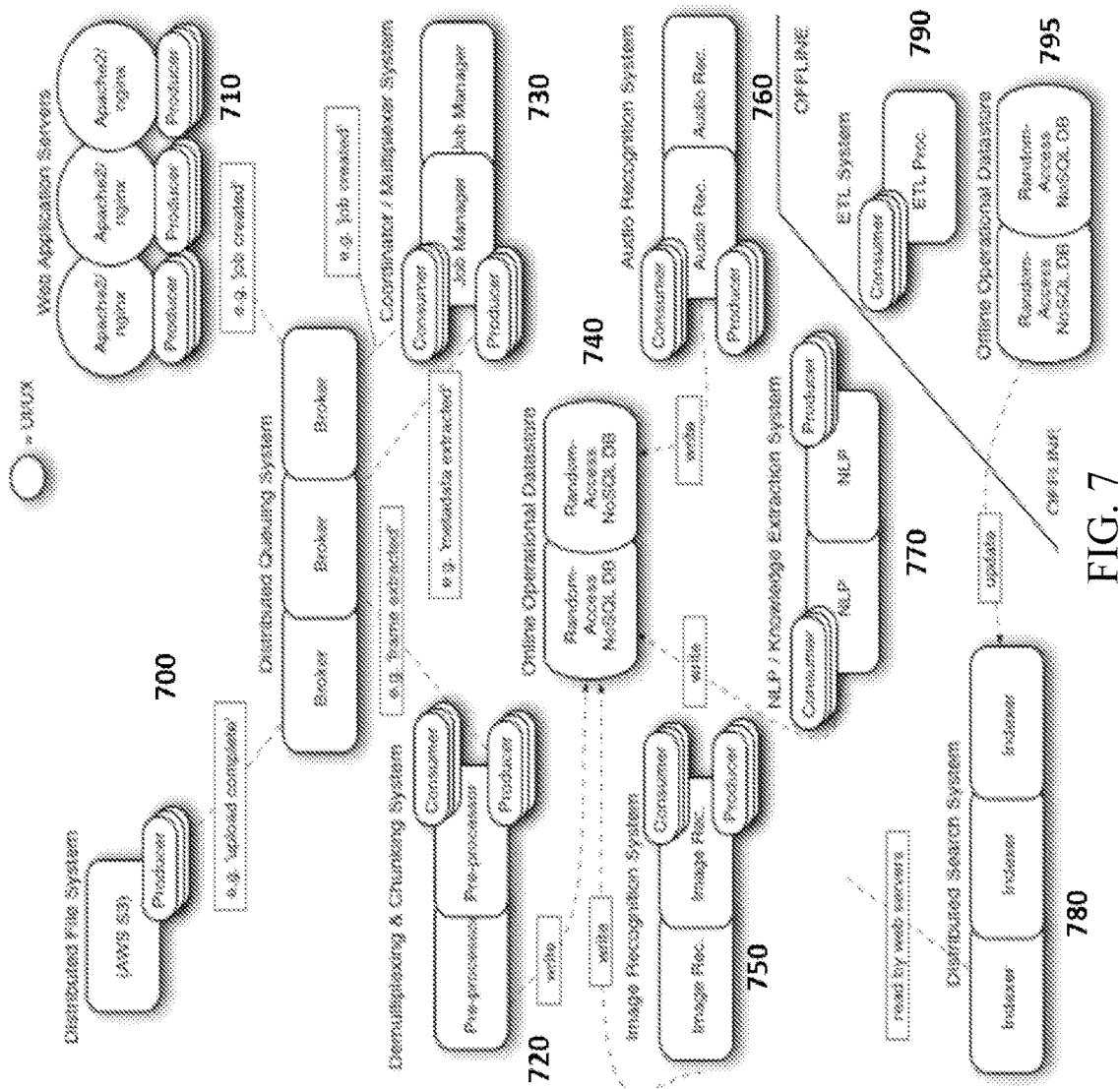


FIG. 7

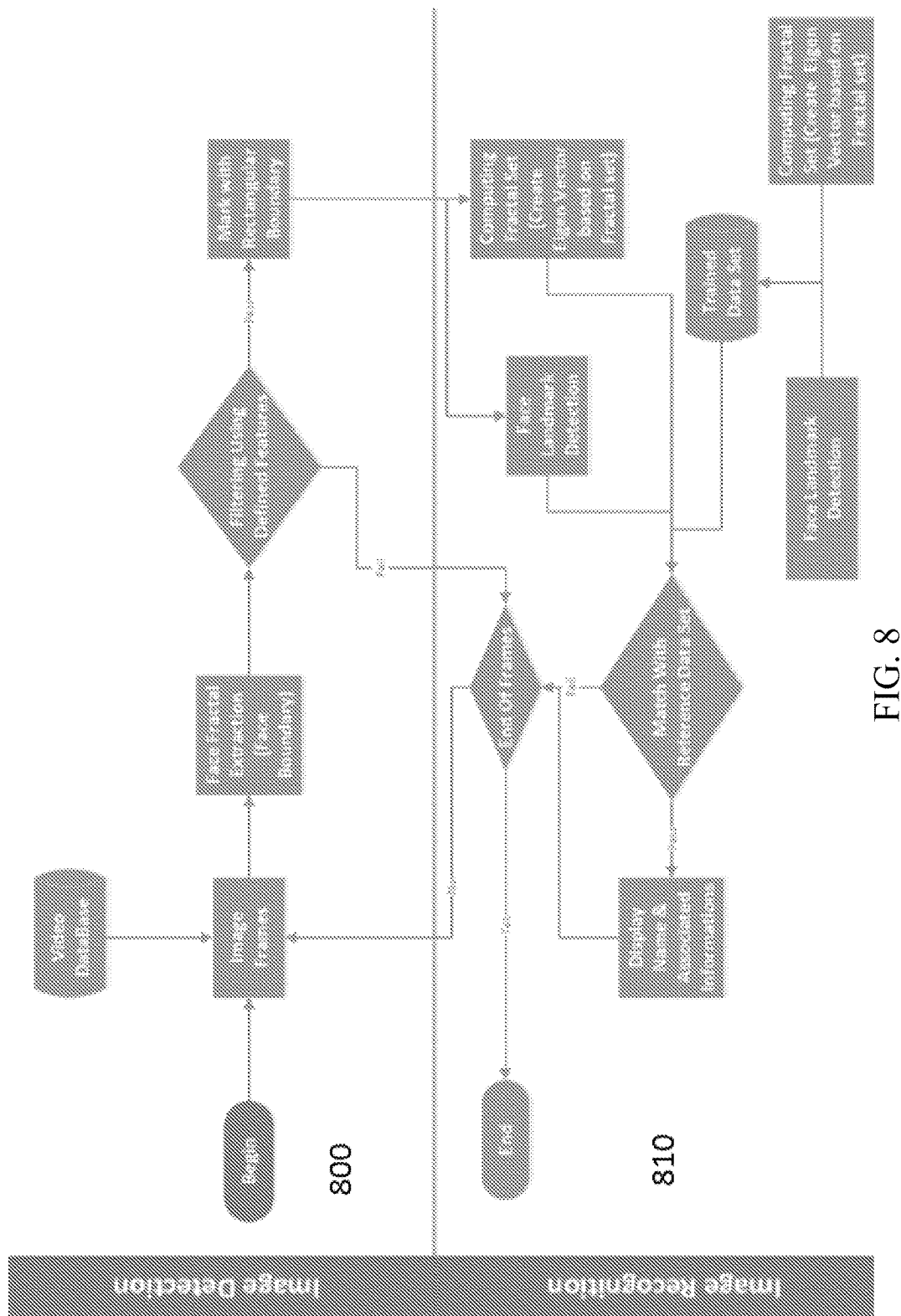


FIG. 8

U.S. Patent

Feb. 26, 2019

Sheet 9 of 9

US 10,218,954 B2



FIG. 9

US 10,218,954 B2

1

VIDEO TO DATA

CLAIM OF PRIORITY

This application claims the benefit under 35 USC 371 to International Application No. PCT/US2015/014940, filed Feb. 7, 2015, which claims priority to U.S. patent application Ser. No. 14/175,741, filed Feb. 7, 2014, which claims priority to U.S. Provisional Patent Application No. 61/866,175, filed on Aug. 15, 2013, and claims priority to U.S. Provisional Patent Application No. 62/021,666, filed Jul. 7, 2014, each of which is incorporated by reference in its entirety.

TECHNICAL FIELD

The present invention relates to a method and a system for generating various and useful data from videos.

BACKGROUND

In the field of image contextualization, distributed reverse image similarity searching can be used to identify images similar to a target image. Reverse image searching can find exactly matching images as well as flipped, cropped, and altered versions of the target image. Distributed reverse image similarity searching can be used to identify symbolic similarity within images. Audio-to-text algorithms can be used to transcribe text from audio. An exemplary application is note-taking software. Audio-to-text, however, lacks semantic and contextual language understanding.

SUMMARY

The present invention is generally directed to a method to generate data from video content, such as text and/or image-related information. A server executing the method can be directed by a program stored on a non-transitory computer-readable medium. The video text can be, for example, a context description of the video.

An aspect of the method can include generating text from an image of the video, converting audio associated with the video to text, extracting topics from the text converted from the audio, cross-referencing the text generated from the image of the video and the topics extracted from audio associated with the video, and generating video text based on a result of the cross-referencing.

In some embodiments, natural language processing can be applied to the generation of text from an image of the video, converting audio associated with the video to text, or both.

In other embodiments, the text from the image of the video can be generated by identifying context, a symbol, a brand, a feature, an object, and/or a topic in the image of the video.

In yet other embodiments, the text from the image can be generated by first segmenting images of the video, and then converting the segments of images to text in parallel. The text from the audio can be generated by first segmenting images of the audio, and then converting the segments of images to text in parallel. The audio can be segmented at spectrum thresholds. The generated text may be of different sizes. The size of the text can be adjusted by a ranking or scoring function that, for example, can adjust the text size based on confidence in the description or relevance to a search inquiry. The text can describe themes, identification of objects or other information of interest.

2

In some embodiments, the method can include generating advertising and/or product or service recommendations based on video content. The video content can be text, context, symbols, brands, features, objects, and/or topics related to or found in the video. An advertisement and/or product or service recommendations can be placed at a specific time in the video based on the video content and/or section symbol of a video image. The advertisement and/or product or service recommendations can also be placed at a specific time as part of the video player, e.g., side panel, and also may be placed on a second screen. In some embodiments, the method can include directing when one or more advertisements can be placed in a predetermined context at a preferred time.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention is further described in the detailed description which follows, in reference to the noted plurality of drawings by way of non-limiting examples of certain embodiments of the present invention, in which like numerals represent like elements throughout the several views of the drawings, and wherein:

FIG. 1 illustrates an embodiment of present invention.

FIG. 2 illustrates an embodiment of image data processing.

FIG. 3 illustrates an embodiment of audio data processing.

FIG. 4 illustrates another embodiment of present invention.

FIG. 5 illustrates various exemplary embodiments of present invention.

FIG. 6 illustrates a flow diagram of an embodiment

FIG. 7 illustrates an embodiment of the architecture of the present invention.

FIG. 8 illustrates a flow diagram of an embodiment of image recognition.

FIG. 9 illustrates an embodiment of a graphical user interface of the present invention.

DETAILED DESCRIPTION

A detailed explanation of the system and method according to exemplary embodiments of the present invention are described below. Exemplary embodiments described, shown, and/or disclosed herein are not intended to limit the claims, but rather, are intended to instruct one of ordinary skill in the art as to various aspects of the invention. Other embodiments can be practiced and/or implemented without departing from the scope and spirit of the claimed invention.

The present invention is generally directed to system, device, and method of generating content from video files, such as text and information relating to context, symbols, brands, features, objects, faces and/or topics found in the images of such videos. In an embodiment, the video-to-content engine can perform the functions directed by programs stored in a computer-readable medium. That is, the embodiments may take the form of a hardware embodiment (including circuits), a software embodiment, or an embodiment combining software and hardware. The present invention can take the form of a computer-program product that includes computer-useable instructions embodied on one or more computer-readable media.

The various video-to-content techniques, methods, and systems described herein can be implemented in part or in whole using computer-based systems and methods. Additionally, computer-based systems and methods can be used

US 10,218,954 B2

3

to augment or enhance the functionality described herein, increase the speed at which the functions can be performed, and provide additional features and aspects as a part of or in addition to those described elsewhere in this document. Various computer-based systems, methods and implementations in accordance with the described technology are presented below.

A video-to-content engine can be embodied by the a general-purpose computer or a server and can have an internal or external memory for storing data and programs such as an operating system (e.g., DOS, Windows 2000™, Windows XP™, Windows NT™, OS/2, UNIX or Linux) and one or more application programs. Examples of application programs include computer programs implementing the techniques described herein for lyric and multimedia customization, authoring applications (e.g., word processing programs, database programs, spreadsheet programs, or graphics programs) capable of generating documents or other electronic content; client applications (e.g., an Internet Service Provider (ISP) client, an e-mail client, or an instant messaging (IM) client) capable of communicating with other computer users, accessing various computer resources, and viewing, creating, or otherwise manipulating electronic content; and browser applications (e.g., Microsoft's Internet Explorer) capable of rendering standard Internet content and other content formatted according to standard protocols such as the Hypertext Transfer Protocol (HTTP). One or more of the application programs can be installed on the internal or external storage of the general-purpose computer. Alternatively, application programs can be externally stored in or performed by one or more device(s) external to the general-purpose computer.

The general-purpose computer or server may include a central processing unit (CPU) for executing instructions in response to commands, and a communication device for sending and receiving data. One example of the communication device can be a modem. Other examples include a transceiver, a communication card, a satellite dish, an antenna, a network adapter, or some other mechanism capable of transmitting and receiving data over a communications link through a wired or wireless data pathway.

The general-purpose computer or server may also include an input/output interface that enables wired or wireless connection to various peripheral devices. In one implementation, a processor-based system of the general-purpose computer can include a main memory, preferably random access memory (RAM), and can also include a secondary memory, which may be a tangible computer-readable medium. The tangible computer-readable medium memory can include, for example, a hard disk drive or a removable storage drive, a flash based storage system or solid-state drive, a floppy disk drive, a magnetic tape drive, an optical disk drive (Blu-Ray, DVD, CD drive), magnetic tape, paper tape, punched cards, standalone RAM disks, Iomega Zip drive, etc. The removable storage drive can read from or write to a removable storage medium. A removable storage medium can include a floppy disk, magnetic tape, optical disk (Blu-Ray disc, DVD, CD) a memory card (Compact-Flash card, Secure Digital card, Memory Stick), paper data storage (punched card, punched tape), etc., which can be removed from the storage drive used to perform read and write operations. As will be appreciated, the removable storage medium can include computer software or data.

In alternative embodiments, the tangible computer-readable medium memory can include other similar means for allowing computer programs or other instructions to be loaded into a computer system. Such means can include, for

4

example, a removable storage unit and an interface. Examples of such can include a program cartridge and cartridge interface (such as the found in video game devices), a removable memory chip (such as an EPROM or flash memory) and associated socket, and other removable storage units and interfaces, which allow software and data to be transferred from the removable storage unit to the computer system.

An embodiment of video-to-content engine operation is illustrated in FIG. 1. At 110, a video stream is presented. The video stream may be in format of (but not limited to): Advanced Video Codec High Definition (AVCHD), Audio Video Interlaced (AVI), Flash Video Format (FLU), Motion Picture Experts Group (MPEG), Windows Media Video (WMV), or Apple QuickTime (MOV), h.264 (MP4).

The engine can extract audio data and image data (e.g. images or frames forming the video) from the video stream.

In some embodiments, the video stream and the extracted image data can be stored in a memory or storage device such as those discussed above. A copy of the extracted image data can be used for processing.

At 120, the video-to-content engine performs an image data processing on the video stream. An example of the image data processing is illustrated in FIG. 2. In FIG. 2, the image data 310 can be segmented into N segments and processed in parallel (e.g., distributed processing 320-1 to 320-N), allowing for near real-time processing.

An example of the video image data processing can be symbol (or object) based. Using image processing technique such as color edge detection, a symbol of a screen or an image of the video can be isolated. The symbol can be identified using an object template database. For example, the symbol includes 4 legs and a tail, and when matched with the object template database, the symbol may be identified as a dog. The object template database can be adaptive and therefore, the performance would improve with usage.

Other image data processing techniques may include image extraction, high-level vision and symbol detection, figure-ground separation, depth and motion perception.

Another example of video image processing can be color segmentation. The colors of an image (e.g., a screen) of the video can be segmented or grouped. The result can be compared to a database using color similarity matching.

Based on the identified symbol, a plurality of instances of the symbol can be compared to a topic database to identify a topic (such as an event). For example, the result may identify the dog (symbol) as running or jumping. The topic database can be adaptive to improve its performance with usage.

Thus, using the processing example above, text describing a symbol of the video and topic relating to the symbol may be generated, as is illustrated in FIG. 9. Data generated from an image and/or from audio transcription can be time stamped, for example, according to when it appeared, was heard, and/or according to the video frame from which it was pulled.

At 330, the engine combines the topics as an array of keys and values with respect to the segments. The engine can segment the topics over a period of time and weight the strength of each topic. Further, the engine applies the topical meta-data to the original full video. The image topics can be stored as topics for the entire video or each image segment. The topic generation process can be repeated for all identifiable symbols in a video in a distributed process. The outcome would be several topical descriptors of the content within a video. An example of the aggregate information that would be derived using the above example would be under-

US 10,218,954 B2

5

standing that the video presented a dog, which was jumping, on the beach, with people, by a resort.

Identifying various objects in an image can be a difficult task. For example, locating (segmenting) and positively identifying an object in a given frame or image can yield false positives—locating but wrongfully identifying an object. Therefore, present embodiments can be utilized to eliminate false positives, for example, by using context. As one example, if the audio soundtrack of a video is an announcer calling a football game, then identification of ball in a given frame as a basketball can be assigned a reduced probability or weighting. As another example of using context, if a given series of image frames from a video is positively or strongly identified as a horse race, then identifying an object to be a mule or donkey can be given a reduced weight.

Using the context or arrangement of certain objects in a given still or static image to aid in computer visual recognition accuracy can be an extremely difficult task given certain challenges associated with partially visible or self-occluded objects, lack of objects, and/or faces, and/or words or an overly cluttered image, etc. However, the linear sequencing of frames from a video—as opposed to a stand-alone image—avails itself to a set images {images x-y} from which context can be derived. This contextual methodology can be viewed as systematic detection of probable image false positives by identifying an object from one video frame (or image) as an anomaly when compared to and associated with a series of image frames both prior and subsequent to the purported anomaly. According to the objects, faces, words, etc. of a given set of frames (however defined), a probability can be associated with an identified anomaly to determine whether an image may be a false positive and, if so, what other likely results should be.

In certain instances, identification of an individual can be a difficult task. For example, facial recognition can become difficult when an individual's face is obstructed by another object like a football, a baseball helmet, a musical instrument, or other obstructions. An advantage of some embodiments described herein can include the ability to identify an individual without identification of the individual's face. Embodiments can use contextual information such as associations of objects, text, and/or other context within an image or video. As one example, a football player scores a touchdown but rather than identifying the player using facial recognition, the player can be identified by object recognition of, for example, the player's team's logo, text recognition of the player's jersey number, and by cross referencing this data with that team's roster (as oppose to another team, which is an example of why the logo recognition can be important). Such embodiments can further learn to identify that player more readily and save his image as data.

Similarly, the audio transcript of a video can be used to derive certain context helpful in identifying and correcting or eliminating image false positives. In this way, an image anomaly or anomalies identified in a given video frame(s) are associated with time (time stamped) and correlated with a time range from the transcribed audio to establish certain probabilities of accuracy.

Moreover, the aforementioned methodologies—establishing probabilities of accuracy of image identification from a set of frames and from the audio transcription—can be combined to improve the results.

In some embodiments, a similar context methodology can be used to identify unknown objects in a given image by narrowing a large, or practically infinite, number of possibilities to a relatively small number of object possibilities

6

and assigning probabilities. For example, neuro-linguistic programming (NLP), neural network programming, or deep neural networks can be utilized to achieve sufficient narrowing and weighting. For further example, based on a contextual review of a large number of objects over a period of time, a series of nodes in parallel and/or in series can be developed by the processor. Upon initial recognition of objects and context, these nodes can assign probabilities to the initial identification of the objection with each node in turn using context and further description to narrow the probabilistic choices of an object. Other methodologies can be utilized to determine and/or utilize context as described herein.

Natural language processing can be useful in creating an intuitive and/or user-friendly computer-human interaction. In some embodiments, the system can select semantics or topics, following certain rules, from a plurality of possible semantics or topics, can give them weight based on strength of context, and/or can do this a distributed environment. The natural language processing can be augmented and/or improved by implementing machine-learning. A large training set of data can be obtained from proprietary or publicly available resources. For example, CBS News maintains a database of segments and episodes of “60-Minutes” with full transcripts, which can be useful for building a training set and for unattended verification of audio segmentation. The machine learning can include ensemble learning based on the concatenation of several classifiers, i.e. cascade classifiers.

At 130, an optional step of natural language processing can be applied to the image text. For example, based on dictionary, grammar, and a knowledge database, the text extracted from video images can be modified as the video-to-content engine selects primary semantics from a plurality of possible semantics. In some embodiments, the system and method can incorporate a Fourier transform of the audio signal. Such filtering can improve silence recognition, which can be useful for determining proper placement of commas and periods in the text file.

In parallel, at 140, the video-to-content engine can perform audio-to-text processing on audio data associated with the video. For example, for a movie video, the associated audio may be the dialog or even background music.

In addition to filtering of the audio signal, images from the video signal can be processed to address, for example, the problem of object noise in a given frame or image. Often images are segmented only to locate and positively identify one or very few main images in the foreground of a given frame. The non-primary or background images are often treated as noise. Nevertheless, these can provide useful information, context and/or branding for two examples. To fine-tune the amount of object noise cluttering a data set, it can be useful to provide a user with an option to dial image detection sensitivity. For certain specific embodiments, identification of only certain clearly identifiable faces or large unobstructed objects or band logos can be required with all other image noise disregarded or filtered, which can require less computational processing and image database referencing, in turn reducing costs. However, it may become necessary or desirable to detect more detail from a frame or set of frames. In such circumstances, the computational thresholds for identification of an object, face, etc. can be altered according to a then stated need or desire for non-primary, background, obstructed and/or grainy type images. Such image identification threshold adjustment capability can be implemented, for example, as user-controlled inter-

US 10,218,954 B2

7

face, dial, slider, or button, which enables the user to make adjustments to suit specific needs or preferences.

An example of the audio data processing is illustrated in FIG. 3. In FIG. 3, the audio data **410** can be segmented into N segments and processed in parallel (e.g., distributed processing **420-1** to **420-N**), allowing for near real-time processing.

In some embodiments, the segmentation can be performed by a fixed period of time. In another example, quiet periods in the audio data can be detected, and the segmentation can be defined by the quiet periods. For example, the audio data can be processed and converted into a spectrum. Locations where the spectrum volatility is below a threshold can be detected and segmented. Such locations can represent silence or low audio activities in the audio data. The quiet periods in the audio data can be ignored, and the processing requirements thereof can be reduced.

Audio data and/or segments of audio data can be stored in, for example, memory or storage device discussed above. Copies of the audio segments can be sent to audio processing.

The audio data for each segment can be translated into text in parallel, for example through distributed computing, which can reduce processing time. Various audio analysis tools and processes can be used, such as audio feature detection and extraction, audio indexing, hashing and searching, semantic analysis, and synthesis.

At **430**, text for a plurality of segments can then be combined. The combination can result in segmented transcripts and/or a full transcript of the audio data. In an embodiment, the topics in each segment can be extracted. When combined, the topics in each segment can be given a different weight.

The audio topics can be stored as topics for the entire video or each audio segment.

At **150**, an optional step of natural language processing can be applied to the text. For example, based on dictionary, grammar, and/or a knowledge database, the text extract from the audio stream of a video can be given context, an applied sentiment, and topical weightings.

At **160**, the topics generated from an image or a frame and the topics extracted from audio can be combined. The text can be cross-referenced, and topics common to both texts would be given additional weights. At **170**, the video-to-content engine generates video text, such as text describing the content of the video, using the result of the combined texts and cross reference. For example, key words indicating topic and semantic that appear in both texts can be selected or emphasized. The output can also include metadata that can be time-stamped with frame references. The metadata can include the number of frames, the range of frames, and/or timestamp references.

FIG. 4 illustrates another embodiment of the present invention. User equipment (UE) **210** can communicate with a server or servers **220** via a network **230**. An exemplary embodiment of the system can be implemented over a cloud computing network.

For exemplary purposes only, and not to limit one or more embodiments herein, FIG. 6 illustrates a flow diagram of an embodiment. A video file is first split into video data and audio data. A data pipeline, indicated in the figure as Video Input/Output, can extract sequences of image frames and can warehouse compressed images in a distributed data store as image frame data. A distributed computation engine can be dedicated to image pre-processing, performing e.g. corner and/or edge detection and/or image segmentation. The engine can also be dedicated to pattern recognition, e.g. face

8

detection and/or logo recognition, and/or other analysis, such as motion tracking. Processed data can be sent to one or more machines that can combine and/or sort results in a time-ordered fashion. Similarly, the Audio Input/Output represents a data pipeline for e.g. audio analysis, compression, and/or warehousing in a distributed file system. The audio can be, for example but not limited to WAV, MP3, or other known formats. Also similarly to the video branch, a distributed computation engine can be dedicated to audio pre-processing, e.g. noise removal and/or volume adjustment, pattern recognition, e.g. transcription and/or keyword detection, and/or other analysis, e.g. identifying unique speakers. Processed audio data can be sent to one or more machines that reassemble transcript segments in their correct time-order. A time-stamped transcript can be sent through an NLP, or other preferred system or analysis, which can transform the data in time-ordered topics and/or subject matter. Both branches converge to output data from parallel video and audio pipelines. The output data can be synced into one or more machines that can combine image and audio generated topics and/or tags which can be applied towards a number of user experiences or user-defined outputs. Such experiences can include search engine optimization, video categorization, recommendation engines, advertisement targeting, content personalization, analytics, etc. The output can include metadata that is time-stamped with frame references. The metadata can include the number of frames, the range of frames, and/or timestamp references.

The UE **210** can include, for example, a laptop, a tablet, a mobile phone, a personal digital assistant (PDA), a keyboard, a display monitor with or without a touch screen input, and an audiovisual input device. In another implementation, the peripheral devices may themselves include the functionality of the general-purpose computer. For example, the mobile phone or the PDA may include computing and networking capabilities and function as a general purpose computer by accessing a network and communicating with other computer systems.

The server **220** can include the general purpose computer discussed above.

The network **230** includes, for example, the Internet, the World Wide Web, WANs, LANs, analog or digital wired and wireless telephone networks (e.g., Public Switched Telephone Network (PSTN), Integrated Services Digital Network (ISDN), and Digital Subscriber Line (xDSL)), radio, television, cable, or satellite systems, and other delivery mechanisms for carrying data. A communications link can include communication pathways that enable communications through one or more networks.

In some embodiments, a video-to-content engine can be embodied in a server or servers **220**. The UE **210**, for example, requests an application relating to the video stream. The servers **220** perform the audio-to-text process on the segmented audio in parallel. The distributed audio-to-text processing reduces the overall response time. This method allows real-time audio-to-text conversion.

The UE **210** communicates with the server **220** via the network **230** for video stream application. The video-to-content engine can generate the video text as illustrated in FIG. 1. The server **220** then generates advertisement (text, images, or animation) based on the video text. In some embodiments, the server adds the advertisement to a specific symbol, image, frame, or a specific time in the video stream. The specific symbol, image, frame, or the specific time in the video stream can be selected based on the video text.

The server **220** can add the audio text to the video stream in real time (i.e. real time close caption).

US 10,218,954 B2

9

The server **220** can generate video recommendation based on a database of the video text. In some embodiments, the server **220** can search videos based on the video text (e.g., via a database of video text). In this fashion, video search can be optimized. Applications for the video search optimization may include search engine optimization (SEO), search engine marketing (SEM), censorship and removal materials of copyright violation.

The video streams can be videos viewed by a user, and the server **220** generates a preference profile for the user using the video data.

In an embodiment, as shown in FIG. **5** for example, a server node can fetch a video file. For example, a URL can be used to fetch the video file from an Internet such as YouTube, and from such URL the video can be scraped. The server can divide the video into chunks of smaller data files for processing on several nodes of a cluster in parallel. For example, the video file can be separated into audio files and image frame files. Each of the types of files can be normalized.

The normalized audio files can be split into constituent files for processing and reduction in parallel by various nodes. Various reduction processes can be performed on the constituent audio files such as phoneme detection and assembly as well as grammar assembly. An output of the audio processing steps can be an extracted text map.

The normalized image frame files can be processed in order to extract various data maps, such as a text map, a tag map, a brand, an object map, a feature map, and/or a tracking map. Such maps can be achieved through various extraction steps. For example, the normalized image frame files can be analyzed for text identification and/or by optical character recognition. The data can be improved through a dictionary verification step. Various maps can be created based on edge detection and/or image segmentation techniques. Such techniques can be improved by focusing on regions of interest, for example based on brands, logos, objects, and/or features of interest. Additionally, or alternatively, pixel gradients of the normalized image frame files can be analyzed and/or the files can be segmented by temporal and/or spatial components, and thus, for example, allow extraction of motion within the video images, which in turn can be used for tracking.

Identification of motion or action in a still image can be a challenge in the vision science field. However, the linear sequencing of frames from a video—as opposed to a stand-alone image—avails itself motion detection. A series of sequential frames can be analyzed in groups to identify actions, rather than merely objects, or, as a manifestation of data, verbs rather than nouns. For example, an object found across several frames can be identified and the object's motion can be determined with a high degree of accuracy. For further example, a processor can analyze a collection of x sequential frames to identify a basketball found in each frame. The processor then can analyze the motion of the basketball to determine that a basket was made by slam-dunking. Or, a certain automobile may be identified in frame a , with a crashed automobile identified in frame z , with the sequential frames in between a and z identifying the action of said car crashing. The accuracy of the action-identification can be improved by utilizing contextual recognition methodologies discussed herein. For example, the probability of positively identifying a slam-dunk action can be increased if Michael Jordan is identified in the context of the video and/or images. Action identification can be further improved, in addition to modeling context of objects, by

10

analyzing human poses, for example by building a learning set of still images capturing known actions.

In an embodiment, as shown in FIG. **7** for example, to generate metadata, several sub-systems may operate on the video. An image recognition sub-system **750** can take frames from the video, and isolate and identify objects present in the frame. An audio recognition sub-system **760** can include automatic speech recognition, sound identification and music identification. A natural language processing (NLP) sub-system **770** can annotate and assign meaning to keywords that are generated by the image and audio sub-systems.

The automatic speech recognition (ASR) model can be a set of statistical structures and operation used for determining words from expected audio signals. The ASR model can consist of an acoustic model (AM) and a language model (LM) between which there is near perfect overlap. The acoustic model can map audio speech features to sounds/word-parts. For example, a series of features might be mapped to the *ah* sound in 'bath.' The language model can consist of a dictionary of known words and their phonetic mappings to sounds/word-parts and a statistical weighting of likelihood of a given word occurring given the previous one or two words. Speech may contain words and phrases not commonly used in "regular" language, e.g., double dribble, free throw, slap shot, high-sticking, etc. Accordingly, the language model can also consist of a topic-specific dictionary, with associated phonetic variants for each words, and may also consist of a statistical N-gram model of word probabilities, e.g., "slam dunk" is a common phrase but will be used more frequently in sports reporting than in general language.

The acoustic model can process audio waveforms to generate a series of speech features based on Mel-Frequency Cepstral Coefficients (MFCCs) that are generated using a series of signal processing techniques including pre-emphasis, cosine or hamming windowing, FFT, Mel-filtering, log power spectrum, DCT and the computation of delta and delta-delta coefficients. The automatic speech recognition can encompass a wide variety of tasks, e.g., connected digit recognition, dialogue systems, large vocabulary continuous speech recognition (LVCSR). Automatic speech recognition can work in part by having a statistical understanding of what words can follow other words.

The following operational sub-systems (**710**, **720**, **730**, **740**, **780**, **790**, **795**) can be utilized to support the metadata generation process. Web application servers **710** can provide clients with the ability to use the service provided by the system, e.g., upload, monitor progress and receive outcomes. A video processing sub-system **720** can transform the video file into data on which the metadata generating sub-systems operate. An auditing/coordination sub-system **730** can monitor the overall system performance, and can generate operational and business analytic data. An operational data storage sub-system **740** can store the generated metadata as well as operational and business analytic data for use in active, online processes. A search sub-system **780** can index client results, and can make them searchable via the web application. An offline data storage system **795** can hold the history of all operations performed in the system including business and operational data. An extract-transform-load (ETL) subsystem **790** can regularly write the to the offline data storage sub-system.

An architecture based on distributed message queuing and distributed data storage **700** may be utilized to build a scalable system, to optimally allocate resources for performance, to enhance failure and overload resiliency. A dis-

US 10,218,954 B2

11

tributed message queuing system may produce data that gets delivered to a particular queue at which time it gets consumed by a component that watches that queue. The distributed queuing system can be removed.

For exemplary purposes only, and not to limit one or more embodiments herein, FIG. 8 illustrates a flow diagram of an embodiment of image recognition. The images can be classified as faces and objects, and typically an image can contain faces and objects.

Image recognition can include two components: image detection 800 and image recognition 810. Image detection can be utilized to determine if there is a pattern or patterns in an image that meet the criteria of a face, image, or text. If the result is positive, the detection processing then moves to recognition. All fractal computations can occur in recognition. Recognition processing can include creating a fractal representation of the face or object that was detected, performing a match to an existing database of faces and objects, and assigning a value (name) to the face or object and then returning to the requesting program.

The system can utilize facial recognition algorithms to identify facial fractals by extracting landmarks from an image of the subject's face. For example, the algorithm may analyze the relative position, size, and/or shape of the eyes, nose, cheekbones, and jaw. These features can then be used to search for other images with matching features. Other algorithms can normalize a gallery of face images and then compress the face data, only saving the fractal data in the image that is useful for face recognition. A probe image can then be compared with the face data. Recognition algorithms can be divided into two main approaches, geometric, which looks at distinguishing features, or photometric, which is a statistical approach that distills an image into values and compares the values with templates to eliminate variances.

The recognition algorithms may include Principal Component Analysis using Eigen faces, Linear Discriminate Analysis, Elastic Bunch Graph Matching using the Fisher-face algorithm, the Hidden Markov model, the Multi-linear Subspace Learning using tensor representation, and the neuronal motivated dynamic link matching. A hybrid using fractal genesis can be constructed to detect the face with elements described above.

Three-dimensional face recognition can also be used. This technique can use 3D sensors to capture fractal information about the shape of a face. This information can then be used to identify distinctive features on the surface of a face, such as the contour of the eye sockets, nose, and chin.

One advantage of 3D facial recognition is that it is not affected by changes in lighting like other techniques. It can also identify a face from a range of viewing angles, including a profile view. Three-dimensional data points from a face vastly improve the precision of facial recognition.

To improve the accuracy of detection, the hybrid can also use the visual details of the skin, as captured in standard digital or scanned images. This technique, called skin texture analysis, turns the unique lines, patterns, and spots apparent in a person's skin into a mathematical fractal space. Tests have shown that with the addition of skin texture analysis, performance in recognizing faces can increase 20 to 25 percent.

The following recognition models may be utilized:

PCA: Derived from Karhunen-Loeve's transformation. Given an s-dimensional vector representation of each face in a training set of images, Principal Component Analysis (PCA) tends to find a t-dimensional subspace whose basis vectors correspond to the maximum variance direction in the original image space. This new subspace is normally

12

lower dimensional ($t \ll s$). If the image elements are considered as random variables, the PCA basis vectors are defined as eigenvectors of the scatter matrix.

Linear Discriminant Analysis (LDA): finds the vectors in the underlying space that best discriminate among classes. For all samples of all classes the between-class scatter matrix SB and the within-class scatter matrix SW are defined. The goal is to maximize SB while minimizing SW, in other words, maximize the ratio $\det(SB)/\det(SW)$. This ratio is maximized when the column vectors of the projection matrix are the eigenvectors of $(SW^{-1} \times SB)$.

Aa eigenspace-based adaptive approach: searches for the best set of projection axes in order to maximize a fitness function, measuring at the same time the classification accuracy and generalization ability of the system. Because the dimension of the solution space of this problem is too big, it is solved using a specific kind of genetic algorithm called Evolutionary Pursuit (EP).

Elastic Bunch Graph Matching (EBGM): All human faces share a similar topological structure. Faces are represented as graphs, with nodes positioned at fiducial points. (eyes, nose . . .) and edges labeled with 2-D distance vectors. Each node contains a set of 40 complex Gabor wavelet coefficients at different scales and orientations (phase, amplitude). They are called "jets". Recognition is based on labeled graphs. A labeled graph is a set of nodes connected by edges, nodes are labeled with jets, edges are labeled with distances.

Kernel Methods: The face manifold in subspace need not be linear. Kernel methods are a generalization of linear methods. Direct non-linear manifold schemes are explored to learn this non-linear manifold.

Trace transform: A generalization of the Radom transform, is a new tool for image processing which can be used for recognizing objects under transformations, e.g. rotation, translation and scaling. To produce the Trace transform one computes a functional along tracing lines of an image. Different Trace transforms can be produced from an image using different trace functional.

3-D Morphable Model: Human face is a surface lying in the 3-D space intrinsically. Therefore the 3-D model should be better for representing faces, especially to handle facial variations, such as pose, illumination etc. Blantz et al. proposed a method based on a 3-D morphable face model that encodes shape and texture in terms of model parameters, and algorithm that recovers these parameters from a single image of a face.

Bayesian Framework: A probabilistic similarity measure based on Bayesian belief that the image intensity differences are characteristic of typical variations in appearance of an individual. Two classes of facial image variations are defined: intrapersonal variations and extrapersonal variations. Similarity among faces is measures using Bayesian rule.

Hidden Markov Models (HMM): These are a set of statistical models used to characterize the statistical properties of a signal. HMM consists of two interrelated processes: (1) an underlying, unobservable Markov chain with a finite number of states, a state transition probability matrix and an initial state probability distribution and (2) a set of probability density functions associated with each state.

Fractal Genesis (Hybrid): The image intensity differences are characteristic of typical variations in appearance of an individual. Human face is a surface lying in the Fractal space intrinsically. Since most of the parameters are self-similar, the Fractal model should be better for repre-

US 10,218,954 B2

13

senting faces, especially to handle facial variations, such as pose, illumination etc. A Hybrid, wherein many proponents of other algorithms are integrated to form a fractal genesis.

An advantage of present embodiments includes the ability to provide real-time or faster-than-real-time content output. This can be achieved through one or more components and/or steps. For example, a video file can be distributed across at least two layers for processing. The audio can be converted to text on at least one layer, and the images can be processed on at least one other layer. In some embodiments, natural language processing can abstract topics, sentiments, temporal topic-tagging, and can be used for further optimization and/or machine learning. The layers can include node clusters for parallel processing chunks of the video file into the preferred content. In some embodiments, the files can be maintained and processed in parallel at each step, and then combined into a single data file as one of the terminal processing steps.

Present embodiments have wide application. For example, video indexing, reverse image lookup, video co-groupings and graph searches, and video similarity indexing, as described herein, can be used for searching, for classification, and for recommendations regarding processed videos. Law enforcement and security industries can implement embodiments for object recognition and motion detection. Media, entertainment, and industrial entities can implement embodiments to monitor for trademark infringement, captioning, advertising and targeting, brand and product monitoring and data collection, and marketing analytics. These exemplary implementation are not intended to be limiting, merely exemplary.

Additionally, or alternatively, to actively fetching and scraping a video, the system and method can be automated as a push system and/or a web crawling system. For example, the server can monitor or online content of specific providers, such as YouTube, Vimeo, the growing myriad of video-content creating websites, or other online video providers. Monitoring of published videos can be tailored to search for extracted data relevant to specific requesters. For example, a purveyor of certain products can be apprised in real-time of new content relevant to the products. Such relevant content can include the context in which the products are found in the video, the appearance of competing products, verification of product placement, and other useful information.

All of the methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the apparatus and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope or the invention. In addition, from the foregoing it will be seen that this invention is one well adapted to attain all the ends and objects set forth above, together with other advantages. It will be understood that certain features and sub-combinations are of utility and may be employed without reference to other features and sub-combinations. This is contemplated and within the scope of the appended claims. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit and scope of the invention as defined by the appended claims.

The invention claimed is:

1. A method to generate video data from a video comprising:

14

generating audio files and image files from the video; distributing the audio files and the image files across a plurality of processors and processing the audio files and the image files in parallel;

converting audio files associated with the video to text; identifying an object in the image files; determining a contextual topic from the image files; assigning a probability of accuracy to the identified object based on the contextual topic;

converting the image files associated with the video to video data, wherein the video data comprises the object, the probability, and the contextual topic; cross-referencing the text and the video data with the video to determine contextual topics;

generating a contextual text, an image, or an animation based on the determined contextual topics;

generating a content-rich video based on the generated text, image, or animation.

2. The method according to claim 1, further comprising: applying natural language processing to the text to determine context associated with the video.

3. The method according to claim 1, further comprising: applying natural language processing to the text to extract a topic.

4. The method according to claim 1, further comprising: processing the image files to extract additional text.

5. The method according to claim 4, wherein the additional text is generated by segmenting the image files before processing the image files in parallel.

6. The method according to claim 1, further comprising: processing the image files to extract an object and determine a motion associated with the object.

7. The method according to claim 1, wherein the text is generated by segmenting the audio before processing the audio files in parallel.

8. The method according to claim 6, wherein the audio files and the image files are segmented at spectrum thresholds.

9. The method according to claim 1, further comprising: generating an advertisement based on the text and the video data.

10. The method according to claim 9, further comprising: placing the advertisement in the video at a preferred time.

11. The method according to claim 1, further comprising: generating metadata.

12. The method according to claim 11, wherein the metadata comprises a number of frames, a range of said frames and/or timestamp references.

13. A system for extracting data from a video, comprising: a computer processor; and

a non-transitory computer readable medium containing instructions directing the system to execute the steps of:

converting audio associated with the video to text; converting images associated with the video to video data,

wherein the video data comprises identified objects; extracting topics from the text and the video data;

assigning probabilities of accuracy for each of the identified objects, wherein the probabilities are based on the extracted topics;

cross-referencing the text, the video data, and the topics with the video to determine contextual topics;

generating a contextual text, an image, or an animation based on the determined contextual topics; and

generate a content-rich video based on the generated contextual text, image, or animation.

14. The system according to claim 11, wherein converting the audio comprises natural language processing.

US 10,218,954 B2

15

15. The system according to claim 11, the computer directs the audio to be converted by at least one node of a cluster and the computer directs the images to be converted by at least one other node of the cluster in parallel.

16. The server according to claim 11, wherein the audio 5 and the images are segmented at spectrum thresholds.

17. The system according to claim 13, wherein the computer directs the system to generate metadata.

18. The system according to claim 13, wherein the metadata comprises a number of frames, a range of said frames 10 and/or timestamp references.

* * * * *

16

CERTIFICATE OF COMPLIANCE

I certify that this brief complies with the type-volume limitation of Fed. R. App. P. 27(d)(2)(A). This brief contains 11,820 words, excluding the parts of the Motion exempted by Fed. R. App. 27(d) and Fed. Cir. R. 27(d). This brief complies with the typeface requirements of Fed. R. App. P. 32(a)(5) and the type style requirements of Fed. R. App. 32(a)(6). This brief has been prepared in a proportionally spaced typeface using Microsoft Word in 14-point Times New Roman font.

Dated: June 26, 2026

Respectfully submitted,

/s/ Elizabeth B. Prelogar
Elizabeth B. Prelogar

Counsel for Appellant Google LLC